Transcription-responsive regulation of *c-myc* proto-oncogene – structural and biophysical studies

Cyprian Daniel Cukier

June 2010

Molecular Structure Division
MRC National Institute for Medical Research
The Ridgeway
Mill Hill, London
NW7 1AA

Division of Biosciences
University College London

This thesis is submitted to University College London for the degree of Doctor of Philosophy

I, Cyprian Daniel Cukier confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

The Far-UpStream Element (FUSE) regulatory system tightly controls the expression of *c-myc* proto-oncogene – a master regulator of cellular proliferation and differentiation. The FUSE mechanism relies on the inter-molecular interactions between a DNA regulatory sequence – the FUSE, a transcriptional activator – FUSE-Binding Protein (FBP) and a transcriptional repressor – FBP-Interacting Repressor (FIR). The FUSE DNA element serves as a sensor of the level of ongoing *c-myc* transcription. The FBP and FIR proteins bind sequentially to the FUSE, first to rapidly increase the expression (FBP) and then to cease it (FIR). Crucial for the regulation is a FBP – FIR interaction that acts as a transcriptional on/off switch.

In my thesis, I used Nuclear Magnetic Resonance (NMR) Spectroscopy to determine a high resolution structure of the first two RNA Recognition Motifs (RRMs) of FIR (FIR RRM1-RRM2) alone and in the complex with an Nbox peptide derived from the FBP protein. The structural details provided explain how a low affinity but specific interaction between FBP and FIR is achieved. Further studies, using NMR and BioLayer Interferometry (BLI), explored the interplay between the three components in the FUSE regulatory system. The results show that FIR RRM1-RRM2 provides independent binding sites for the FUSE and FBP protein and that both FBP and FIR proteins are able to bind simultaneously to the DNA. Furthermore, FBP protein tethers FIR to the DNA, thereby contributing to the control of the transcriptional switch. In summary, these data extend our understanding of the *c-myc* proto-oncogene regulation and suggest a strategy for controlling *c-myc* expression in cancerous cells.

Acknowledgments

The work presented in this thesis would not have been performed without the help from many people to whom I express my gratitude.

First and foremost, my supervisor Dr Andres Ramos for giving me a freedom to explore the world of science, but also guidance when it was necessary. Thank you, Andres, for your constant support, belief, understanding and patience, and of course, for giving me an opportunity to do PhD in your laboratory!

David Hollingworth, a heart of the lab, for explaining and showing me many laboratory techniques and for always being there to answer difficult and simple questions when I did not know the answer. Dave also performed some cloning and protein purification that is a part of the work described in this thesis. Thank you, Dave!

I would like to acknowledge Dr Stefan Kindler for a gift of the plasmid encoding rSiahBP, which was used as a template to create FIR constructs.

Dr Irene Diaz-Moreno and Dr Maria-Flor Garcia-Mayoral, the former post-doc fellows in the lab, for a warm welcome in the lab and a lot of help with computing, NMR and laboratory work. Irene also performed SIA analysis of FIR RRM1-RRM2 that is described in this thesis. Thank you, Irene and Flor!

Dr Adela M. Candel and Dr Giuseppe Nicastro, the current post-doc fellows in the lab, for being the great people to discuss the issues of biophysics, NMR and structure calculations. Thank you! I would like to also thank Adela for the coffee meetings that were a great break from the lab work.

People from MRC Biomedical NMR Centre, Dr Thomas A. Frenkiel, Dr Geoff Kelly and Dr Alain Oregioni, for the constant help in running NMR experiments. Geoff was also very helpful in the whole process of structure determination.

Dr Stephen R. Martin, Dr Stephen J. Smerdon and Dr Peter B. Rosenthal, the members of my thesis committee, for the regular meetings, discussions and suggestions that contributed to the work presented here. Stephen R. Martin helped me also to perform and analyse CD and BLI experiments that are described here and he critically read the manuscript of this thesis.

Dr Cesira de Chiara for discussions about ARIA software and John McCormick for a never-ending supply of TEV protease!

Silvia Kralovicova and Dr Sunil Prasannan, members of the Ramos' lab, and all people at the National Institute for Medical Research who I had opportunity to meet for the very friendly atmosphere!

I would like to thank my girlfriend, Jola, for her constant encouragement, belief and patience over the past four years. I had, I have and I will have wonderful life with You! Thank You also for all your help and suggestions during writing this thesis.

My parents, for bringing me up and supporting my education – without them I would not have been where I am. Dziękuję Wam, Mamo i Tato!

All my PhD fellows: Josquin, Peter, Beatris, Goli, Olga, Maria and Kate for the great relaxing events – beer, pool and table tennis meetings ;)

My friends from studies in Kraków: Justyna, Olga, Magda, Jarek i Kuba for the regular meetings with a lot of fun! I hope it will continue!

Table of contents

Table of figures

Table of tables

Table of appendices

Abbreviations

1D, 2D, 3D	one-, two- and three-dimensional
3’-UTR	3’ untranslated region
ABP	ARE-binding protein
ARE	adenine- uracil-rich element
ARIA	ambiguous restraints for iterative assignment
AU	arbitrary unit
bHLHLZ	basic-helix-loop-helix/leucine zipper
BLI	biolayer interferometry
bp, kbp	base pair, kilo base pair
BSA	bovine serum albumin
c-myc	cellular homologue of *myc*
CD	circular dichroism
cDNA	complementary DNA
CNS	crystallography and NMR system
CRD	coding region determinant
CRD-BP	CRD-binding protein
CSP	chemical shift perturbation
Da, kDa	dalton, kilodalton
ds	double-stranded
EC	elongation complex
EMSA	electrophoretic mobility shift assay
FBP	FUSE-binding protein
FIR	FBP-interacting repressor
FUSE	far-upstream element
GST	glutathione-S-transferase
HAT	histone acetyltransferase
H-bond	hydrogen bond
HDAC	histone deacetylase
hetNOE	$\{^{1}H\}$–^{15}N heteronuclear NOE
HisTag	hexa-histidine tag
HMQC	heteronuclear multi-quantum coherence

HPLC	high performance liquid chromatography
HSQC	heteronuclear single-quantum coherence
IMAC	immobilised metal ion affinity chromatography
Inr	initiator element
iPS	induced pluripotent stem cells
IPTG	isopropyl β-D-1-thiogalactopyranoside
K_d	dissociation constant
KH	K-homology motif
LB	Luria-Bertani
L-myc	small cell lung cancer-related *myc*
MAD	MAX dimerisation protein
MAX	MYC-associated factor-X
miRNA	microRNA
mRNA	messenger RNA
myc	a gene responsible for myelocytomatosis
N-myc	neuroblastoma-related *myc*
NMR	nuclear magnetic resonance
NOE	nuclear Overhauser effect
NOESY	NOE spectroscopy
OD_{600}	optical density at 600 nm
PPC	pre-promoter escape complex
ppm	parts per million
RMSD	root mean square deviation
RRM	RNA recognition motif
rRNA	ribosomal RNA
SDS-PAGE	sodium dodecyl sulphate polyacrylamide gel electrophoresis
SEC	size exclusion chromatography
SEC-LS/RI/UV	size exclusion chromatography coupled with light scattering, refractive index and ultraviolet absorbance
SIA	scaffold-independent analysis
siRNA	small interfering RNA
S/N	signal-to-noise ratio
ss	single-stranded

TALOS	torsion angle likelihood obtained from shifts and sequence similarities
τ_c	overall rotational correlation time
TCEP	tris(2-carboxyethyl) phosphine
TEV	tobacco etch virus
TFA	trifluoroacetic acid
TOCSY	total correlated spectroscopy
tRNA	transfer RNA
UHM	U2AF homology motif
ULM	UHM ligand motif
UV	ultraviolet

Chapter 1

Introduction

1.1 Physiological role of the c-MYC protein

The *myc* oncogene was identified in 1978 as a transforming sequence of the M29 avian tumour virus, an infectious agent of myelocytomatosis (*myc*) (Sheiness et al., 1978). Soon afterwards cellular homologues of *myc* (*c-myc*) were discovered in birds (Sheiness and Bishop, 1979) and in humans (Collins and Groudine, 1982). In the past 30 years the transforming potential of *c-myc* has led to a sustained effort to understand its role in cell growth and proliferation, and in cancer. However, *c-myc* biology is very complex and many questions still remain unanswered (Meyer and Penn, 2008).

c-MYC is a member of the human MYC protein family (together with L-MYC (small cell lung cancer-related MYC (Nau et al., 1985)) and N-MYC (neuroblastoma-related MYC (Schwab et al., 1983)). The protein has a broad spectrum of functions – it stimulates cell cycle progression, enables cell growth, blocks differentiation, induces apoptosis, activates angiogenesis, induces genome instability and causes transformation (Figure 1A) (Oster et al., 2002). c-MYC regulates transcription of 10-15% of human genes comprising transcription factors, mRNA metabolism proteins, cell cycle proteins, mitochondrial proteins, DNA repair factors, telomerases and cytokines (Dang et al., 2006; Eilers and Eisenman, 2008; Levens, 2002; Wierstra and Alves, 2008). It also affects transcription of many non-coding RNAs, including tRNAs, rRNAs and miRNAs (Kenneth and White, 2009). Recently c-MYC has been implicated in a direct (transcription-independent) regulation of global chromatin structure, translation and DNA replication (Figure 1B) (Cole and Cowling, 2008; Herold et al., 2009; Knoepfler, 2007).

c-MYC functions as an activator and as a repressor and its effect on the expression of any single gene is relatively modest (Knoepfler, 2007; Oster et al., 2002). The activation function of c-MYC is linked to the MYC/MAX/MAD system (Wierstra and Alves, 2008). c-MYC and MAX dimerisation protein (MAD) are short lived and they regulate target genes antagonistically (Wierstra and Alves, 2008). Both c-MYC and MAD proteins heterodimerise with a ubiquitous and stable MYC-associated factor-X (MAX) via a common basic-helix-loop-helix/leucine zipper (bHLHLZ) dimerisation domain and, as a heterodimer, bind to E-box (5'-CACGTG-3') sequences in the promoters of regulated genes (Nair and Burley, 2006; Oster et al., 2002; Wierstra and Alves, 2008). The MYC/MAX heterodimer can activate expression

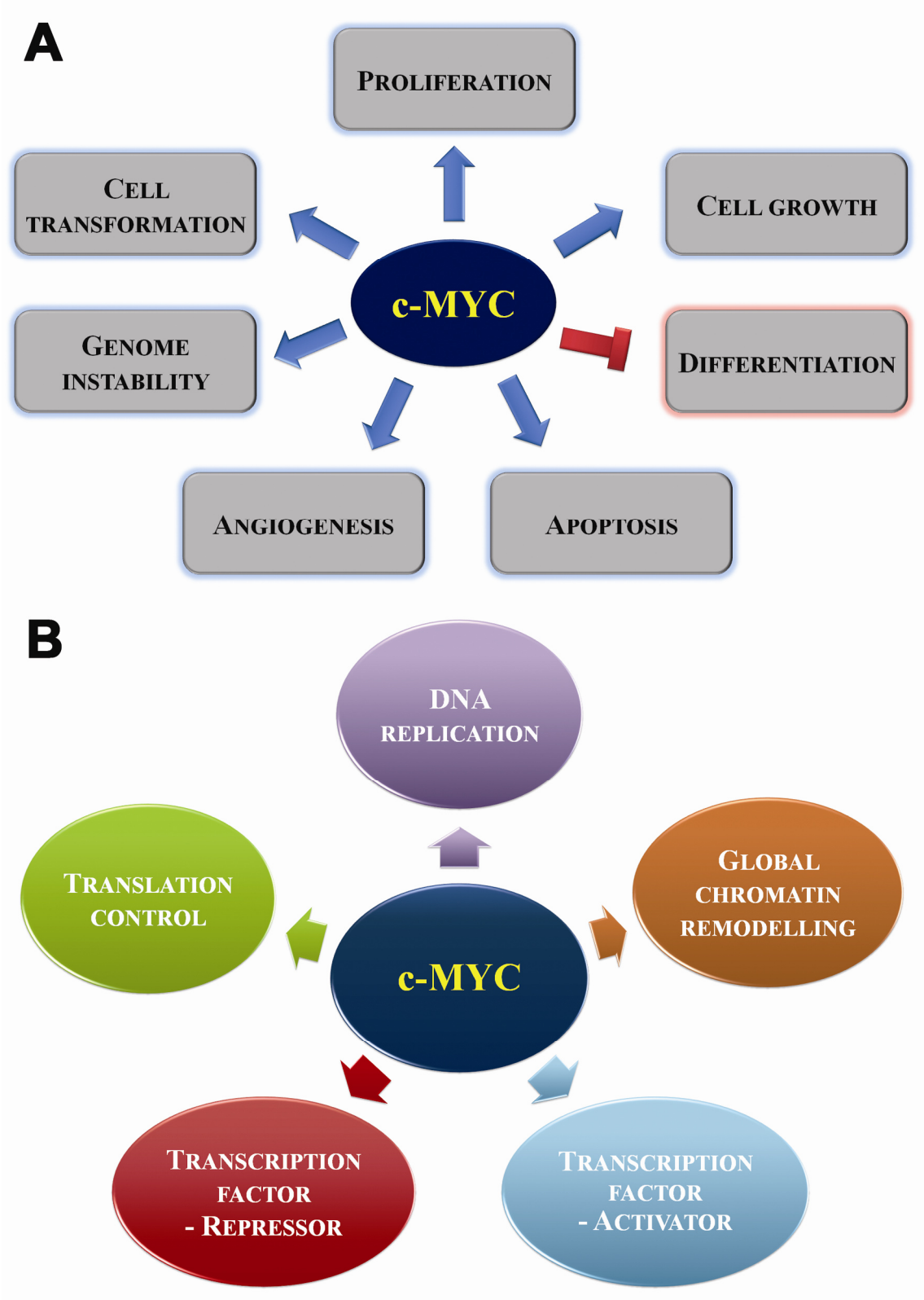

Figure 1 c-MYC has multiple functions in the cell. A) A summary of processes controlled by c-MYC. B) The different modes of action of c-MYC.

via at least three different mechanisms: 1) recruitment of histone acetyltransferase (HAT) complexes to the promoter, 2) recruitment of ATP-dependent chromatin remodelling complexes to the promoter and 3) induction of promoter clearance of the RNA polymerase II at the target promoter (Oster et al., 2002). The MAD/MAX heterodimer represses expression of target genes by recruiting histone deacetylase (HDAC) complexes to the promoter (Wierstra and Alves, 2008).

c-MYC-mediated repression does not involve direct binding of the protein to DNA (Oster et al., 2002). Instead, c-MYC interferes with the function of different activatory factors that act via initiator elements (Inr) (e.g. YY1, TFII-I, Miz-1) or other DNA regulatory sequences (e.g. NF-Y, Sp1/Sp3) present in the regulated promoters (Oster et al., 2002).

The global chromatin remodelling function of c-MYC is based on the recognition of target sequences in the DNA and a balance between c-MYC/MAX heterodimers, which recruit HAT complexes, and MAX/MAD heterodimers, which recruit HDAC complexes (Knoepfler, 2007). Unlike in transcriptional regulation the target DNA sequences are both genic and intergenic and this allows extensive modification of the balance between euchromatin and heterochromatin (Knoepfler, 2007).

c-MYC also increases translation of specific mRNAs (e.g. cyclin T1, CDK9) by enhancing recruitment of cap RNA methyltransferase and promoting 5' mRNA guanine methylation, which is essential for protein-coding gene expression (Cole and Cowling, 2008).

Finally, c-MYC localises to the early sites of DNA replication and interacts with numerous components of the pre-replicative complex, probably playing a role in origin selection (Cole and Cowling, 2008; Herold et al., 2009).

1.2 c-MYC and cancer

Because c-MYC has so many important physiological roles the deregulation of *c-myc* expression is likely to be detrimental for the cell (Vita and Henriksson, 2006). Indeed, since its discovery c-MYC has been shown to be a very strong activator of

carcinogenesis and has been linked to a broad range of cancer pathologies, including Burkitt's and non-Burkitt's lymphomas, breast cancer, prostate cancer, gastrointestinal cancer, melanoma, multiple myeloma and myeloid leukaemia (Nesbit et al., 1999; Oster et al., 2002; Vita and Henriksson, 2006). The vital role of c-MYC as a regulator of the cell's destiny is emphasised by the fact that it is one of the four transcription factors that together reprogram somatic cells to induced pluripotent stem (iPS) cells (Eilers and Eisenman, 2008; Takahashi and Yamanaka, 2006). Although c-MYC is not necessary for the generation of iPS, it substantially increases the efficiency of this process (Nakagawa et al., 2008).

In spite of the clear transforming potential of c-MYC initial studies failed to identify mutations in the coding sequence of the gene in transformed cells (Meyer and Penn, 2008). However, subsequent studies linked c-MYC function in tumorigenesis to three types of large genome rearrangements at the *c-myc* locus: 1) insertional mutagenesis, 2) chromosomal translocation, and 3) amplification (Meyer and Penn, 2008). Insertional mutagenesis relies on insertion of viral sequences into the *c-myc* locus that enhance gene expression (Meyer and Penn, 2008). Chromosomal translocations result in juxtaposition of the *c-myc* gene to potent activatory elements, as in Burkitt's lymphoma where an immunoglobulin enhancer drives a high level of *c-myc* production (Meyer and Penn, 2008; Nesbit et al., 1999; Vita and Henriksson, 2006; Wierstra and Alves, 2008). Finally, gene amplification increases gene dosage in the cell and leads to increased c-MYC activity (Meyer and Penn, 2008; Nesbit et al., 1999).

Further studies have shown that c-MYC can be deregulated not only by genome rearrangements but also by a broad spectrum of direct and indirect mechanisms that target its expression and/or activity, including an activated upstream signalling cascade, point mutations within coding and non-coding *c-myc* sequences, enhanced translation and increased protein stability (Meyer and Penn, 2008; Nesbit et al., 1999; Oster et al., 2002; Vita and Henriksson, 2006). Two well studied examples are point mutation of T58 in c-MYC (observed in Burkitt's lymphomas) (Oster et al., 2002) and increased stability of β-catenin, a co-activator for the TCF-4 transcription factor (observed in gastrointestinal cancers) (Nesbit et al., 1999; Oster et al., 2002). Mutation of T58 abolishes phosphorylation at this residue – this leads to inefficient ubiquitination and stabilisation of the c-MYC protein (Oster et al., 2002). β-catenin is usually sequestered in the cytoplasm by the APC protein, which promotes its degradation (Nesbit et al.,

1999). Mutations that disrupt protein–protein interactions (either in β-catenin or in APC) have been found in cancers with increased c-MYC production (Nesbit et al., 1999). The stabilised β-catenin enters the nucleus and enhances a function of the TCF-4 transcription factor that targets the *c-myc* proto-oncogene (Nesbit et al., 1999).

No single function of c-MYC has been correlated with its transforming potential and it is generally accepted that tumorigenesis is a result of global deregulation of the cell cycle, differentiation, cell growth, genome instability and angiogenesis (Meyer and Penn, 2008). Importantly, c-MYC also activates apoptosis, which serves as a safety valve to eliminate the cells with misregulated *c-myc* (Meyer and Penn, 2008; Wierstra and Alves, 2008). In cancerous cells *c-myc* deregulation is always accompanied by disruptions of the apoptotic pathways that immortalise the cells (Meyer and Penn, 2008). Concurrently, resurrecting apoptotic mechanisms in c-MYC-driven tumours leads to increased mortality of transformed cells (Goga et al., 2007).

Various anti-cancer strategies have been developed to target c-MYC either by decreasing its level (e.g. by blockage of *c-myc* expression at the DNA, mRNA or protein level, or by promotion of c-MYC protein degradation) or by modulating c-MYC function (e.g. by disruption of the c-MYC/MAX interaction, or by inhibition of the expression of c-MYC target genes) (Vita and Henriksson, 2006). However, only recently a direct evidence that c-MYC inhibition could be an efficient therapy was demonstrated in a mouse model (Soucek et al., 2008).

1.3 Regulation of *c-myc* expression

c-myc expression is strictly correlated with cellular proliferation (Figure 2) (Wierstra and Alves, 2008). In quiescent cells c-MYC is practically absent – one transcript and 500 protein molecules per resting fibroblast (Chung and Levens, 2005; Wierstra and Alves, 2008). Upon mitogen stimulation the protein is immediately and robustly produced and this drives the cell into the G1 phase (Wierstra and Alves, 2008). The peak of expression (measured by the mRNA level) is observed 2 h post-stimulation and results in a 10- to 40-fold increase in the amount of protein that quickly drops to a lower level that is kept steady in proliferating cells – as long as growth factors are

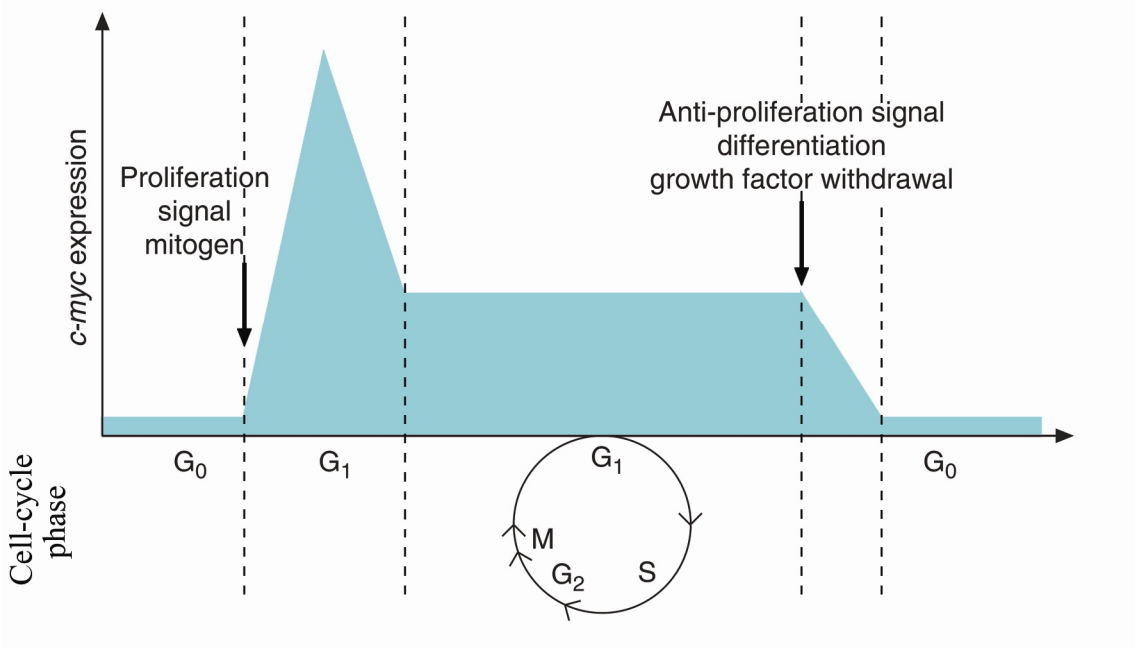

Figure 2 *c-myc* expression throughout the cell cycle (adapted from Wierstra and Alves, 2008).

present (Liu et al., 2006; Wierstra and Alves, 2008). Anti-proliferation, differentiation or starvation signalling results in a rapid decline in *c-myc* expression (Wierstra and Alves, 2008).

The rapid regulation of c-MYC concentration in response to signalling relies on a fast turnover of the protein in the cell (Chung and Levens, 2005; Wierstra and Alves, 2008). Low abundance and short half-lives (20-30 min.) of c-MYC mRNA and protein are important factors that allow a rapid switch in the activation of the *c-myc* (Chung and Levens, 2005; Wierstra and Alves, 2008). Furthermore, the multiple levels at which *c-myc* expression is regulated potentiate the possibilities for efficient up- and down-regulation during the cell cycle (Wierstra and Alves, 2008). However, at the same time the proto-oncogene has to be tightly controlled to protect the cells from the deleterious effect of its over-expression (Chung and Levens, 2005). The necessity to combine these two conflicting requirements makes the regulatory network of *c-myc* expression very complex (Chung and Levens, 2005; Meyer and Penn, 2008; Wierstra and Alves, 2008).

The sophisticated post-transcriptional regulation of *c-myc* ensures rapid clearance of c-MYC mRNA and protein in the absence of pro-proliferating signalling (Chung and Levens, 2005; Wierstra and Alves, 2008). The c-MYC mRNA has two independent instability determinants: 1) the coding region determinant (CRD) located in the sequence corresponding to the C-terminal domain of the protein and 2) the adenine-uracil-rich element (ARE) located in the 3' untranslated region (3'-UTR) (Meyer and Penn, 2008). CRD is specifically recognised by a CRD-binding protein (CRD-BP) that protects c-MYC mRNA against endonucleolytic cleavage (Bernstein et al., 1992; Prokipcak et al., 1994; Sparanese and Lee, 2007). Normally CRD-BP is expressed only in foetal tissues but it has also been detected in cancerous cells (Ioannidis et al., 2003). AREs are recruiting elements for ARE-binding proteins (ABPs), *trans*-acting factors that regulate the rate of *c-myc* mRNA decay via deadenylation, decapping and then the exosome-mediated 3'-5' degradation pathway (Barreau et al., 2005).

The fast turnover of c-MYC protein relies on proteasomal degradation and is driven by at least two ubiquitin ligases that act in phosphorylation dependent (FBW7) and independent (SKP2) ways (Amati, 2004; Eilers and Eisenman, 2008). The two ligases recognise different regions of c-MYC and have different biological roles – as tumour suppressor (FBW7) and as proto-oncogene (SKP2) (Amati, 2004).

1.4 Complexity of *c-myc* proto-oncogene transcription

c-myc proto-oncogene is located in the human chromosome 8 (Wierstra and Alves, 2008). The locus is extensively regulated via chromatin remodelling and the methylation and acetylation state of histones correlates with *c-myc* transcription activity and cellular proliferation (Wierstra and Alves, 2008). The gene has four distinct transcription start sites (P0, P1, P2 and P3) and two polyadenylation sites, which are utilised with different frequencies (Wierstra and Alves, 2008). The role of these alternative transcripts is not clear, but it has been observed, for example, that the ratio of P1 : P2 transcription start sites usage is increased from 1 : 10 – 1 : 5 in normal cells to 1 : 1 and higher in abnormal cells (Wierstra and Alves, 2008).

No single regulatory pathway is alone responsible for the physiological activation of the *c-myc* promoter. Instead, the promoter is responsive to multiple inputs from over 30 different transcription factors that bind to distinct *cis* regulatory sequences spread over a DNA region some 3 kbp in length (Wierstra and Alves, 2008). These transcription factors are the effectors of different signalling pathways which are in turn activated by a range of extracellular signals including growth factors (e.g. PDGF, EGF), mitogens (e.g. LPS, anti-CD3), hormones (e.g. glucocorticoids, testosterone), vitamins (e.g. vitamin A and D) and cytokines (e.g. TNF, IL-1) (Wierstra and Alves, 2008). Furthermore, the c-MYC protein directly and indirectly affects the *c-myc* expression level (Wierstra and Alves, 2008). The direct regulation is a typical negative feedback loop where c-MYC dimerises with MAX and interferes with activation driven by Inr-binding and E2F transcription factors (Wierstra and Alves, 2008). The indirect mechanisms can provide both positive and negative inputs to the *c-myc* promoter because c-MYC acts as an activator/repressor of its own activators/repressors (Wierstra and Alves, 2008).

The large number of inputs of similar importance minimises transcriptional noise and therefore the risk of unwanted *c-myc* expression (Chung and Levens, 2005). This guarantees tight control of *c-myc*, but could hinder effective (rapid) activation as many different transcription factors would have to act concurrently to significantly increase the level of transcription (Chung and Levens, 2005). A rapid response is possible because 1) there is a paused polymerase II complex at the promoter that simply has to be released, saving the time that would be required for assembly of the whole

complex, and 2) an additional regulatory mechanism, called FUSE–FBP–FIR, provides a positive feedback loop regulation (Chung and Levens, 2005; Liu et al., 2006; Nechaev and Adelman, 2008; Wierstra and Alves, 2008).

1.5 The FUSE–FBP–FIR regulatory system

The FUSE–FBP–FIR system is designed to assure a rapid regulation of *c-myc* transcription in proliferating cells, accurately shaping the peak of *c-myc* expression (Chung and Levens, 2005; Liu et al., 2006; Wierstra and Alves, 2008). The FUSE mechanism rapidly up- and down-regulates the expression of *c-myc* in response to its own transcription and therefore *c-myc* transcription regulation may be considered as a two-step regulation (Liu et al., 2006; Wierstra and Alves, 2008). First, in response to multiple inputs at the promoter a decision is made to initiate *c-myc* expression by increasing the escape of the paused polymerase (Liu et al., 2006; Wierstra and Alves, 2008). Then the regulation of the transcription is taken over by the FUSE–FBP–FIR mechanism that further drives the expression of the gene (Liu et al., 2006; Wierstra and Alves, 2008).

The FUSE mechanism is based on a *cis* DNA regulatory element called far-upstream element (FUSE) and two *trans*-acting proteins – the FUSE-binding protein (FBP) and the FBP-interacting repressor (FIR) (Liu et al., 2006). The FUSE is an AT-rich element located 1.7 kbp upstream of the *c-myc* P2 transcription start site (Avigan et al., 1990). The actively transcribing RNA polymerase II is one of the most potent generators of torsional stress and introduces an overtwisting in the DNA helix downstream of the polymerase (positive torsional stress) and an undertwisting in the DNA helix upstream of the polymerase (negative torsional stress) (Kouzine and Levens, 2007). The force generated upstream of the promoter during transcription is sufficient to melt the double-stranded FUSE DNA regulatory element that is indeed found in a single-stranded conformation in proliferating cells (Bazar et al., 1995; Duncan et al., 1994; He et al., 2000; Kouzine et al., 2004; Kouzine et al., 2008; Michelotti et al., 1996). The FBP and FIR proteins both bind to the single non-coding strand of the FUSE (ssFUSE) and act respectively as an activator and a repressor

(Avigan et al., 1990; Duncan et al., 1994; He et al., 2000; Liu et al., 2000; Liu et al., 2006; Michelotti et al., 1996). The FUSE DNA works as a sensor that detects a low level of *c-myc* transcription initiated by multiple inputs into the promoter and switches on the FUSE–FBP–FIR regulatory system to rapidly up-regulate *c-myc* expression (Liu et al., 2006).

Liu *et al.* have proposed a model of the FUSE–FBP–FIR mechanism that can be summarised as three steps: 1) initiation of *c-myc* transcription in response to environmental signals that leads to the melting of FUSE; 2) a boost in transcription caused by the FBP protein and 3) shut off of the transcription by the FIR protein (Figure 3) (Liu et al., 2006). In resting cells the *c-myc* proto-oncogene is silent: the FUSE element is masked by a nucleosome and there is a paused pre-promoter escape complex at the P2 transcription start site (Liu et al., 2006; Michelotti et al., 1996). Upon activation, which can be triggered by many concurrently and/or sequentially acting factors, the rate of escape of the paused polymerase is augmented and this leads to an increase in transcription (Liu et al., 2006). Simultaneously, chromatin remodelling proteins are recruited to unmask the FUSE element (Liu et al., 2006; Michelotti et al., 1996). The ongoing transcription causes an accumulation of negative torsional stress upstream of the *c-myc* proto-oncogene, which melts the FUSE DNA and makes it single-stranded (Kouzine et al., 2004; Kouzine et al., 2008; Liu et al., 2006). Only then is the FBP protein recruited – it interacts with transcription factor TFIIH and establishes a loop with the transcriptional machinery (He et al., 2000; Liu et al., 2001; Liu et al., 2006). FBP increases the 3'-5' helicase activity of the XPB subunit of TFIIH boosting *c-myc* transcription (Liu et al., 2001; Liu et al., 2006). FBP-mediated stimulation may be a more complex process as TFIIH acts at multiple steps during transcription initiation and elongation (Weber et al., 2005). The increase in productive transcription mediated by FBP binding to TFIIH introduces additional twists into the DNA, increasing the negative torsional stress and melting a larger section of the FUSE (Liu et al., 2006). Subsequently, the FIR protein is recruited to the FBP–FUSE complex (Liu et al., 2006). FIR also interacts with the XPB subunit of TFIIH and suppresses its 3'-5' helicase activity, slowing down the transcription (Liu et al., 2000; Liu et al., 2006). The torsional stress is gradually dissipated by topoisomerases and this leads to re-annealing of the FUSE sequence and dissociation of the FBP (Liu et al., 2006). Dissociation of the activator leads to a further decrease of transcription, ejection of FIR and complete re-annealing of the FUSE (Liu et al., 2006). The FUSE again

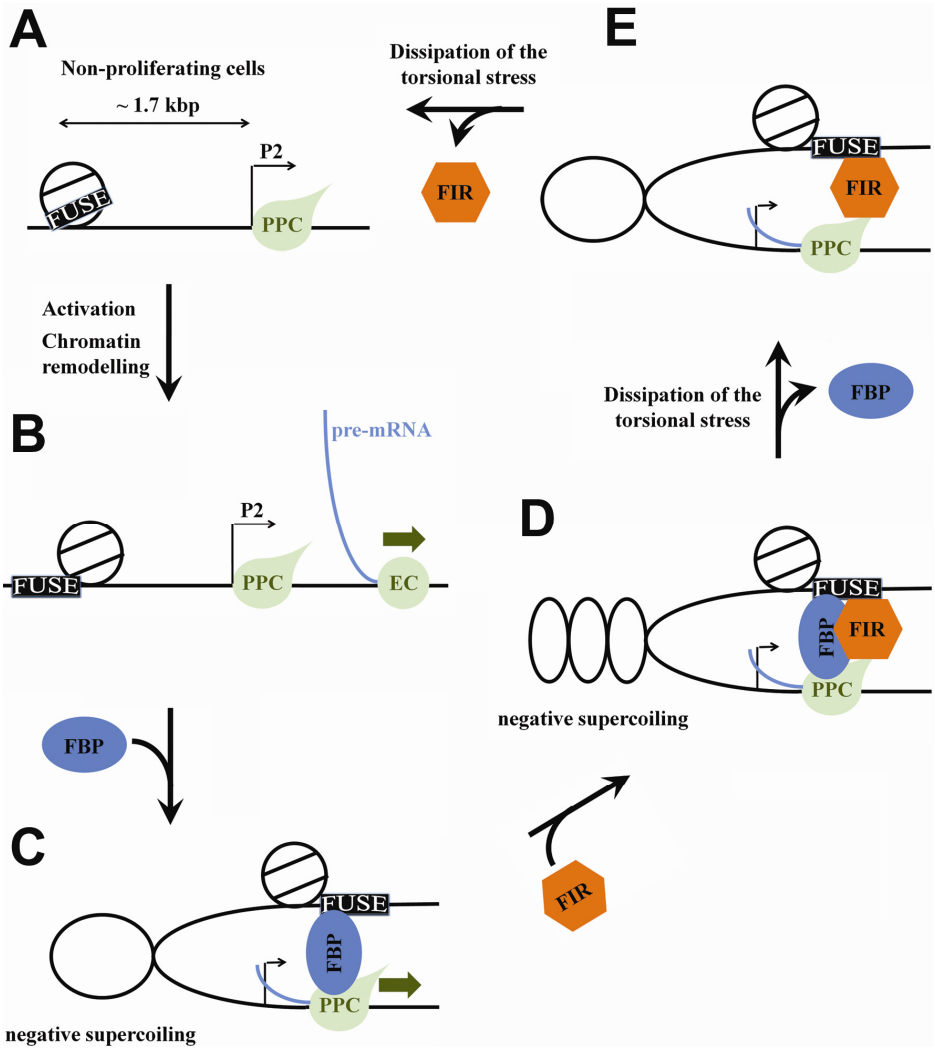

Figure 3 Proposed model of FUSE–FBP–FIR regulation of *c-myc* proto-oncogene (adapted from Liu et al., 2006). A) In non-proliferating cells, the FUSE is masked by a nucleosome while a pre-promoter escape complex (PPC) is paused at the P2 *c-myc* promoter. B) A stimulating event converts PPC into an elongation complex (EC) and causes chromatin remodelling, which unmasks the FUSE. C) The resulting transcription creates negative supercoiling upstream of *c-myc* that melts the FUSE element, which then recruits the FBP. FBP contacts the TFIIH at the *c-myc* promoter and boosts transcription to create a peak in expression. D) As transcription proceeds the topologically closed FBP–TFIIH loop accumulates negative supercoiling leading to further melting of the FUSE. This causes FIR to be recruited via interactions with the ssFUSE DNA and FBP and TFIIH proteins. E) FIR represses transcription, decreasing the dynamic stress. This leads to FBP ejection. A) The low level of transcription causes dissociation of the polymerase from the PPC complex. Finally, without activation, the torsional stress is dissipated, FIR dissociates and the *c-myc* promoter becomes silent.

becomes masked by a nucleosome and the promoter is again silent (Liu et al., 2006). This complex mechanism allows the rapid up- and down-regulation of *c-myc* expression, creating a peak in the mRNA level that reaches a maximum at ~ 2 h post-stimulation and ends within the following ~ 4 h (Liu et al., 2006). This model is supported by two observations: 1) a mutation in TFIIH that abolishes the creation of the loop between FUSE and the transcription start site changes the expression profile of *c-myc* and 2) the same mechanism can be employed to reprogram expression from an unrelated promoter if a FUSE element is inserted in the appropriate place (Liu et al., 2006).

1.6 The FBP family

The human FBP protein family comprises the FBP, FBP2 and FBP3 proteins (Davis-Smyth et al., 1996). All three proteins were initially identified in connection with the FUSE DNA regulatory mechanism and were shown to bind a non-coding strand of FUSE and to activate gene transcription (Davis-Smyth et al., 1996; Duncan et al., 1994). FBP proteins share a very similar domain organisation with three functionally distinct segments: an N-terminal domain, a central nucleic acid binding domain and a C-terminal activation domain (Figure 4A) (Davis-Smyth et al., 1996; Duncan et al., 1994). The N-termini of FBP and FBP2 are involved in transcription repression (Chung et al., 2006; Duncan et al., 1996). Originally this activity was mapped to the first 107 residues of FBP but further analysis highlighted the importance of a short sequence (Nbox) within this region, which was proposed to form an amphipathic α-helix (Chung et al., 2006; Duncan et al., 1996). Interestingly the Nbox sequence of FBP3, which does not repress transcription, is the most divergent of the three sequences (Chung et al., 2006). The central parts of the FBPs contain four K-homology motifs (KH motifs), a fold originally identified as an RNA recognition domain in the hnRNP K protein and later found to be common in proteins engaged in ssDNA and ssRNA recognition (Davis-Smyth et al., 1996; Duncan et al., 1994; Siomi et al., 1993; Valverde et al., 2008). The C-termini of the FBP proteins contain 2-4 repeats

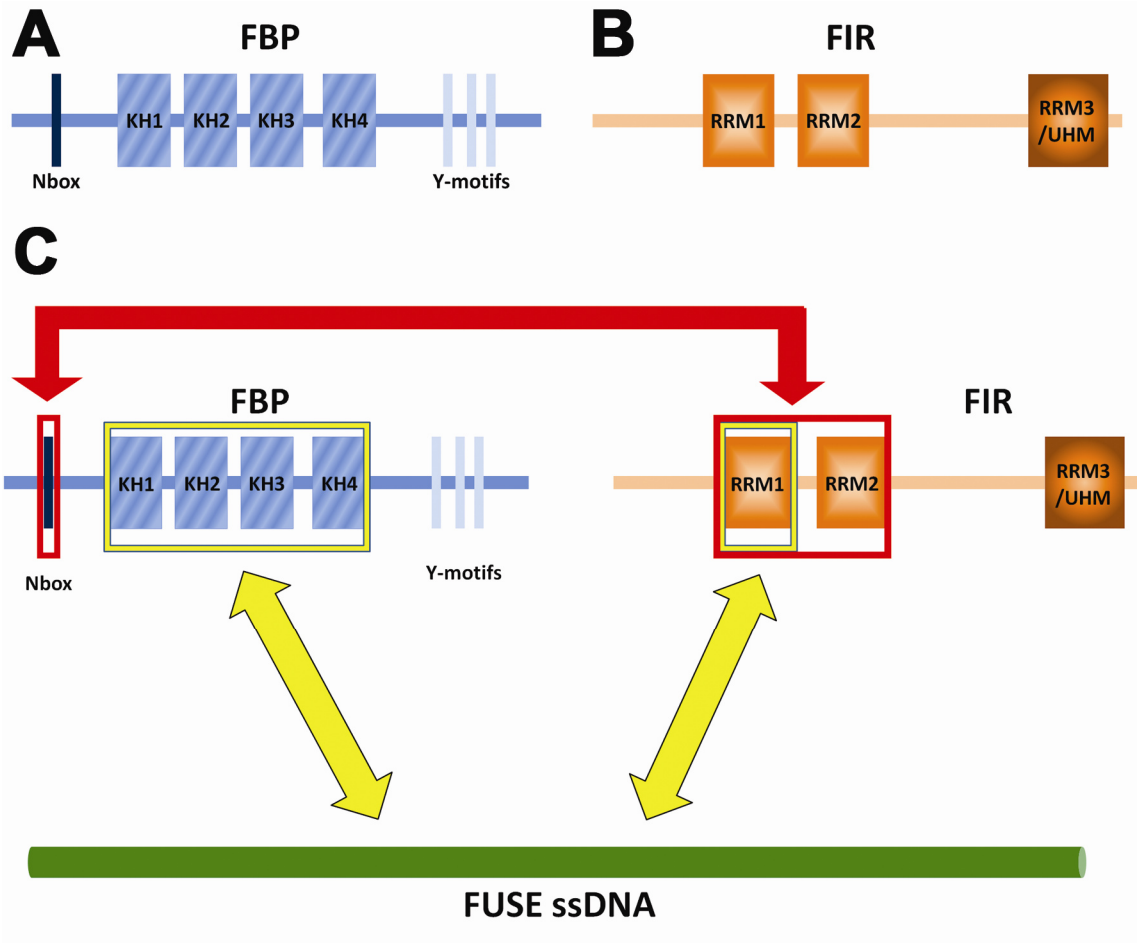

Figure 4 Domain organisation of the FBP and FIR proteins and a summary of inter-molecular interactions in the FUSE–FBP–FIR system. A) Domain organisation of FBP protein. FBP2 and FBP3 are very similar but they contain respectively four and two Y-motifs instead of the three found in FBP. B) FIR comprises three RRM domains. C) Schematic representation of inter-molecular contacts between the three components of the system. The regions of proteins involved in protein–nucleic acid and protein–protein interactions are boxed in yellow and red respectively.

of Y-motifs (AW(A/E)(A/E)YY) that are engaged in transcriptional enhancement (Davis-Smyth et al., 1996; Duncan et al., 1996).

In spite of the similar domain organisation of the FBP proteins, siRNA experiments revealed that they regulate only partially overlapped subsets of genes and therefore are not functionally redundant (Chung et al., 2006). The diverse roles played by the proteins are further emphasised by their different distribution in different cell types – the FBP proteins are expressed in a broad range of tissues – and by additional functions performed by only some of the family members (Davis-Smyth et al., 1996; Wang et al., 1998). The FBP2 protein and its homologues have been implicated in splicing, miRNA processing, cellular localisation and mRNA degradation (Briata et al., 2005; Gherzi et al., 2004; Kroll et al., 2002; Min et al., 1997; Rehbein et al., 2002; Trabucchi et al., 2009), whilst the FBP protein has been found in large RNA-containing complexes (Rothe et al., 2006). The FBP proteins occupy the FUSE regulatory sequence of the *c-myc* proto-oncogene at different times during transcription regulation, but the individual functions of each protein have not been studied in detail (Chung et al., 2006). Interestingly overexpression of FBP and FBP3 has been correlated with high *c-myc* expression and high proliferation in renal cancer (Weber et al., 2008).

1.7 The FIR protein

FIR acts both as a transcriptional repressor of the *c-myc* proto-oncogene and as a splicing factor (Liu et al., 2000; Page-McCaw et al., 1999). The protein is encoded by a gene with two alternatively spliced exons and the absence of one of them has been connected with tumorigenesis (Matsushita et al., 2006). FIR comprises three RNA recognition motifs (RRM) – a common ssDNA and ssRNA binding module (Figure 4B) (Clery et al., 2008). The first two RRMs (RRM1 and RRM2) are canonical nucleic acid binding domains that display the highly conserved aromatic residues necessary for DNA and RNA binding (Clery et al., 2008). The third RRM domain (RRM3) belongs to the U2AF homology motif (UHM) RRM protein-interacting subfamily and interacts with several protein partners containing short peptide UHM ligand motifs (ULM) (Corsini et al., 2009; Kielkopf et al., 2004).

The FIR protein is localised mainly in the nucleus and is expressed in different amounts depending on the cell and tissue type (Liu et al., 2000; Page-McCaw et al., 1999). The transcriptional repression activity of FIR has been mapped to the N-terminal, unstructured region of the protein, that is involved in a direct interaction with TFIIH (Liu et al., 2000). It has recently been shown that expression of FIR by an adenovirus vector has an anti-tumour effect on human cancer cells both *in vitro* and *in vivo* (Matsushita et al., 2009).

1.8 Interactions between FUSE, FBP and FIR

The comprehensive understanding of FUSE-mediated regulation of *c-myc* requires full characterisation of inter-molecular contacts in the FUSE–FBP–FIR system and currently available data are summarised in this section.

The FBP and FIR FUSE binding activity was originally mapped to the KH domains of FBP and to the first two RRM domains of FIR (Figure 4C) (Chung et al., 2006; Davis-Smyth et al., 1996; Duncan et al., 1994). Furthermore, the solution structure of KH3 and KH4 of FBP bound to a FUSE-derived sequence revealed that the ssDNA is recognised in an extended conformation and interacts with the canonical nucleic acid binding surfaces of the KH domains (Braddock et al., 2002). FBP binds preferentially to T- and G-rich sequences but this preference is not stringent and the binding is context-dependent (Benjamin et al., 2008; Braddock et al., 2002). An X-ray structure of the first two RRM domains (RRM1-RRM2) of FIR bound to a FUSE-derived sequence showed that the ssDNA recognition occurs via the β-sheet surface of RRM1, which is a canonical nucleic acid binding surface of an RRM domain (Clery et al., 2008; Crichlow et al., 2008). Interestingly RRM2 of FIR is not involved in the interaction, despite possessing conserved aromatic residues usually engaged in nucleic acid recognition (Clery et al., 2008; Crichlow et al., 2008). Instead, the β-sheet surface of RRM2 interacts with α-helices of RRM1 to form a hydrophobic inter-domain interface (Crichlow et al., 2008). Furthermore, it was found that two FIR RRM1-RRM2 molecules dimerise upon binding to ssDNA (Crichlow et al., 2008). The dimerisation surface spans both RRMs and is favoured due to a fixed inter-domain orientation

between RRM1 and RRM2 within a monomer (Crichlow et al., 2008). The protein–protein interaction is weak because it occurs only after the two FIR RRM1-RRM2 molecules are bound to the ssDNA and the dissociation constants (K_d) for the two binding sites on the FUSE-derived sequence are ~ 2 μM and ~ 75 μM (Crichlow et al., 2008). Although FIR RRM1-RRM2 was crystallised in the complex with FUSE-derived ssDNA only single nucleotides (C or A) were modelled in both binding sites of the dimer due to a poor electron density (Crichlow et al., 2008). These two identified nucleotides are not in agreement with the T-rich sequence preference of FIR observed in missing-base interference analysis (Benjamin et al., 2008).

Protein–protein interactions between FBP and FIR were studied with the yeast-two hybrid assay (Chung et al., 2006). These experiments showed that RRM1-RRM2 of FIR interact with a construct of FBP that contains an Nbox sequence and KH domains (Chung et al., 2006). Mutational analysis of the Nbox and chimeric constructs of different FBP proteins suggested that the Nbox is the determinant for FIR RRM1-RRM2 interaction with FBP and that several hydrophobic residues in this putative α-helix are crucial for the binding (Figure 4C) (Chung et al., 2006). KH domains of FBP contribute to FIR–FBP interaction in the presence of FUSE-derived ssDNA because they bind to the ssDNA molecule, the same molecule FIR binds to, increasing the local concentration of the Nbox (Chung et al., 2006). Furthermore, the inter-protein recognition may modify the position of the two proteins on the FUSE element (Benjamin et al., 2008).

FBP and FIR both interact with the XPB subunit of TFIIH that is important for the transcriptional regulatory function of both proteins (Liu et al., 2001; Liu et al., 2000). Immunoprecipitation and pull-down assays mapped these interactions to the N-terminal repression domain of FIR and to the C-terminal activation domain of FBP (Liu et al., 2001; Liu et al., 2000).

1.9 Rationale for the project

The molecular interactions within the FUSE–FBP–FIR system have been only partially characterised. Although a qualitative description of the interactions and

information on the structure of several components are available (and discussed above) a coherent description of the system from a structural and biophysical point of view is missing. Lack of this information severely limits the molecular understanding of FUSE-mediated regulation and therefore our ability to interfere with it.

The FUSE mechanism is based on a single activator (FBP) and a single repressor (FIR). Central to FUSE regulation is the FIR–FBP interaction that mediates FIR recruitment and transcriptional shut-off, thereby defining the length and size of the peak of *c-myc* transcription. Therefore structural and quantitative characterisation of the RRM1-RRM2 of FIR and its interactions with FBP and the non-coding strand of FUSE (henceforth referred to as ssFUSE) is of particular importance. Furthermore, to accurately describe models for FBP and FIR recruitment to and ejection from ssFUSE it is necessary to quantify the strength of the FIR–FUSE interaction and to compare it with the FBP–FUSE interaction.

The information available on the FUSE–FBP–FIR system suggests that the complex undergoes significant intra- and inter-molecular motions to manage the appropriate contacts both in space and in time. To dissect this flexible and dynamic mechanism we need to provide a structural description of different, often weak complexes. Nuclear magnetic resonance (NMR) spectroscopy is one of two techniques (the other one being X-ray crystallography) that allows one to determine the structure of macromolecules with atomic resolution. Importantly, NMR spectroscopy has specific features that make it particularly useful for the study of the mechanism of FUSE-mediated *c-myc* regulation. Among them are: the possibility of studying the sample in solution, the ability to study flexibility, conformational changes and inter-molecular interactions within the same experimental framework and the incremental nature of the information obtained during an NMR study that allows one to obtain data on folding states and binding affinities prior to the availability of a high resolution structure. Moreover, previous studies on the identification of protein and DNA regions involved in the interactions (see section 1.8 Interactions between FUSE, FBP and FIR) allow me to reduce the system under study to a size amenable to NMR spectroscopy. The information obtained can be subsequently expanded to larger and more complex systems with the help of a complementary technique.

Biolayer interferometry (BLI) is a useful method for the study of interactions between large molecules because its optical-based detection does not impose an upper limit on the size of the system being studied. On the contrary, the larger the molecules

being investigated, the larger is the shift of the interference signal (the BLI observable). BLI allows one to test different experimental set-ups quickly and with a small amount of material (typically nM to μM concentrations in 200 μl wells). Furthermore, the ability to use low concentrations expands the measurable K_d range towards tighter interactions compared to the range accessible by NMR. This is because dissociation constants can be accurately measured in direct assays (in contrast to competition ones) only at concentrations of studied molecules that are comparable to the K_d and the sensitivity limitations of NMR impose a $K_d > \sim 1$ μM limit on studied complexes.

In summary, NMR and BLI complement each other in the study of macromolecular interactions in general and in the particular for the FUSE–FBP–FIR system.

Misregulation of *c-myc* expression has been associated with many cancers and it has been shown that *c-myc* can drive carcinogenesis (Nesbit et al., 1999; Oster et al., 2002; Vita and Henriksson, 2006). Further studies proved that *c-myc* is a potential target in an anti-cancer therapy (Soucek et al., 2008; Vita and Henriksson, 2006). Nevertheless the complexity of *c-myc* regulation hinders development of efficient and safe therapies. FUSE–FBP–FIR is a very attractive option for medical treatment as it is able to rapidly stop the *c-myc* transcription initiated by different signalling pathways (Wierstra and Alves, 2008). However, a full understanding of this transcription-responsive mechanism is a necessary prerequisite to design the treatment strategies based on manipulation of the FUSE–FBP–FIR system.

The aim of this thesis is to provide a more accurate molecular picture of the FUSE–FBP–FIR system. It focuses on the switch between activation by FBP protein and repression by FIR protein that is central to the *c-myc* transcription regulation. The work includes the high resolution structure of the first two RRM domains of FIR (FIR RRM1-RRM2), the complex between FIR RRM1-RRM2 and an amphipathic α-helix of FBP protein (Nbox) and the qualitative and quantitative description of the interactions in the FUSE–FBP–FIR system.

Chapter 2

Materials and Methods

2.1 Protein structure determination by NMR

This section focuses on the well-established methodologies of protein structure determination by NMR that I have used during my thesis.

NMR spectroscopy can be employed to study the structure, dynamics and interactions of biological macromolecules. In NMR, magnetically active nuclei provide information on their local chemical environment and therefore on the structure and motions of the molecule. A major limitation of NMR is its sensitivity to the size of the system to be studied – with standard methodologies ~ 50 – 100 kDa molecules/ complexes represent a challenge. However, the possibility of studying the systems in solution, of performing time-resolved analysis and of investigating dynamic systems justifies the importance of NMR in the study of biologically relevant questions.

Standard NMR high resolution protein structure determination can be summarised in four steps (Linge et al., 2001; Markley et al., 2003) (Figure 5). First, sequential assignment of the protein backbone resonances is achieved by analysing a number of 3-dimensional (3D) NMR through-bond experiments which correlate intra-residual and sequential ^{15}N, $^{1}H^{N}$, $^{13}C^{\alpha}$, $^{13}C^{\beta}$ and $^{13}C'$ resonances. Then, the amino acid spin systems are assigned to identify the remaining ^{13}C, ^{1}H and ^{15}N frequencies. Subsequently, distance and angle restraints are obtained from NMR observables (i.e. nuclear Overhauser effect (NOE) correlations, exchange experiments, *J* couplings, chemical shift patterns and residual dipolar couplings). Finally, these restraints are inserted into a molecular dynamics protocol to calculate a family of structures that satisfy both the chemical geometry and the provided NMR restraints.

2.1.1 Backbone assignment

The first step towards structure determination is an assignment of different nuclear magnetic resonances observed in the spectra to a specific nucleus of the protein. Initially, the $^{1}H^{N}$, ^{15}N, $^{13}C^{\alpha}$, $^{13}C'$ backbone and $^{13}C^{\beta}$ resonances (Figure 6A) of each residue are identified and sequentially connected using triple-resonance experiments. In the recorded spectra, $^{1}H^{N}$ and ^{15}N resonances of a residue are correlated to the $^{13}C^{\alpha}$,

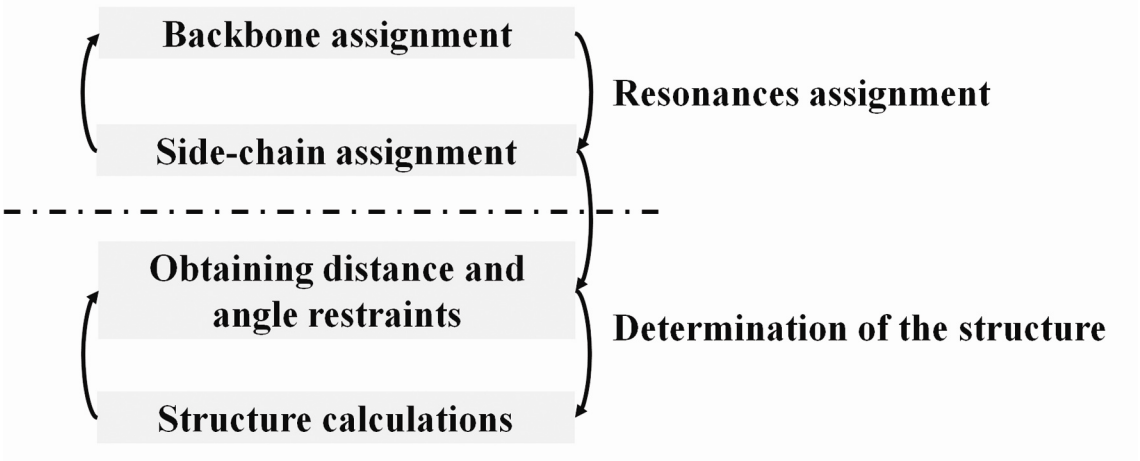

Figure 5 Flow-chart of a standard NMR structure determination protocol.

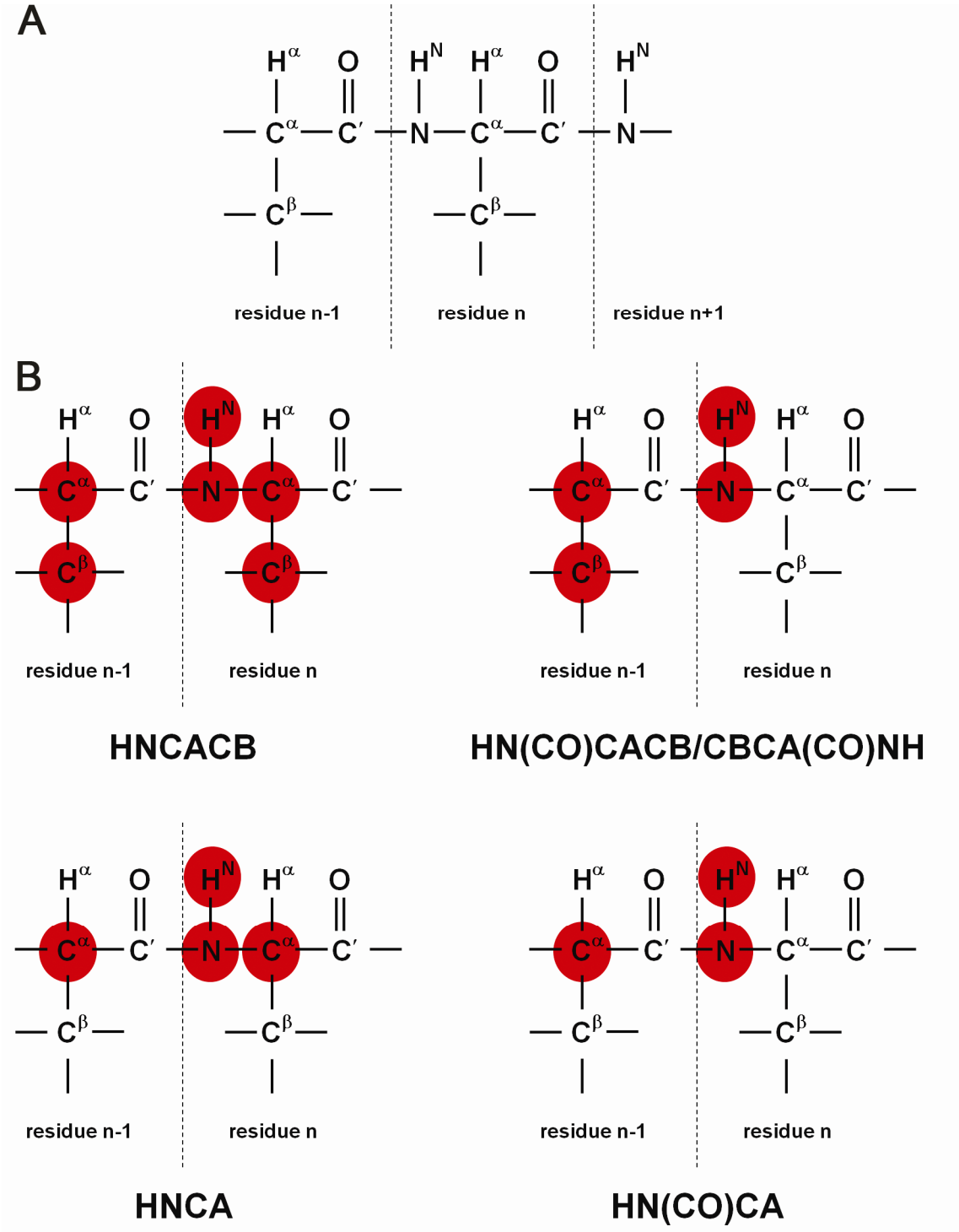

Figure 6 Backbone NMR experiments used in this study. A) Nomenclature. B) Backbone experiments used in my thesis for the assignment. The atoms correlated by the different backbone NMR experiments are marked with red circles. Dashed lines separate the individual residues.

$^{13}C^{\beta}$ and/or $^{13}C'$ resonances of the same and/or preceding residue. The most information-dense experiment for backbone assignment is the 3D HNCACB (Figure 6B). However, this experiment is also the most sensitive to an increase of the molecular weight. The data from the other 3D NMR backbone experiments (HN(CO)CACB, CBCA(CO)NH, HNCA and HN(CO)CA) (Figure 6B) complement the HNCACB data and fill the gaps existing for the fastest relaxing and the overlapped resonances. Where the information obtained from backbone experiments is ambiguous or incomplete, through-space 3D NOE spectroscopy (NOESY) experiments, that are less sensitive to molecular weight and connect a wider range of resonances, can help to identify the missing sequential connections (Cavanagh, 2007). The stretches of sequentially assigned resonances are then matched to unique positions within the primary protein sequence. The process relies on the characteristic distribution of ^{13}C chemical shifts in peptidic chains, that restrict the possibilities of matching with the protein sequence (particularly distinctive are $^{13}C^{\alpha}$ of glycines and $^{13}C^{\beta}$ of alanines, serines and threonines) (Wishart et al., 1995).

For a small well-behaved protein, triple-resonance backbone NMR experiments make sequential assignment of spin systems straightforward. However, the performance of these experiments is strongly dependent on the size and aggregation state of the investigated molecules. In the case of FIR RRM1-RRM2, a moderate molecular weight (~ 20 – 25 kDa) and a transient aggregation experienced at higher concentrations of the protein (> 0.5 mM) significantly affected signal-to-noise (S/N) ratio of the recorded spectra and hindered the assignment process. The NOESY experiments turned out to be an invaluable help to resolve any ambiguities and complete some missing assignments. Nevertheless, not all resonances were identified and unambiguously assigned, especially in loop regions, where internal mobility of the protein may lead to exchange broadening of the resonances and where dispersion of observed chemical shifts is low.

2.1.2 Side-chain assignment

Side-chain assignment is routinely achieved by combining 3D through-bond and through-space NMR experiments (Cavanagh, 2007). The former are 3D HCCH total correlated spectroscopy (TOCSY) and 3D ^{15}N TOCSY experiments that correlate all ^{13}C-attached intra-residual protons to each other or to backbone amides, respectively. The latter are 3D ^{15}N NOESY and 3D ^{13}C NOESY experiments that correlate an individual ^{1}H to other ^{15}N- (^{15}N NOESY) or ^{13}C-attached (^{13}C NOESY) protons within ~ 6 Å distance. The sequential assignment of backbone resonances facilitates the assignment of side-chains, because amino acid spin system correlations can be found starting from the backbone resonances and then the assigned chemical shifts can be validated against the tabulated values for each amino acid type (Wishart et al., 1995). In theory, the through-bond experiments could provide a complete side-chain assignment, as they correlate all protons and heteronuclei (^{13}C and ^{15}N) within the protein. However, there are two difficulties when working on larger proteins (> 10 kDa). Firstly, the increased overall rotational correlation time (τ_c) results in broad peaks and a rapid relaxation of the magnetisation. To increase the sensitivity shorter TOCSY mixing periods are used that often eliminate the correlations between distant ^{1}H in the side-chains (e.g. K $^{1}H^{N}$ – $^{1}H^{\varepsilon 2/\varepsilon 3}$, I $^{1}H^{\alpha}$ – $Q^{\delta 1}$). Secondly, the larger proteins have more residues of the same type increasing the spectra overlap. This is especially apparent in the methylene and methyl regions of the spectra. Fortunately, the information obtained from TOCSY experiments can be completed and validated with through-space NOESY experiments, which rely on the dipolar couplings between the two spins and are not as sensitive as through-bond scalar coupling interactions to the increase of the τ_c of the molecule. In the NOESY spectra, both intra- and inter-residual ^{1}H-^{1}H correlations are observed for nuclei separated by up to ~ 6 Å, depending on the relaxation properties of the resonances and the mixing times used. Importantly, at short mixing time the strongest cross-peaks are typically observed for intra-residual protons and a coherent pattern of correlations allows identification of individual spin systems. Regardless of the approach taken, the best measure of the validity of the assignment results is the self-consistency of all of the information derived from the chemical shift values, through-bond scalar coupling interactions and through-space dipolar coupling interactions. In the case of FIR RRM1-RRM2, the main source of

information for the side-chain assignment were the NOESY spectra, because of the long τ_c of FIR RRM1-RRM2 both in isolation and in the complex with the FBP Nbox peptide.

2.1.3 NMR-derived structural restraints

2.1.3.1 NOE distance restraints

The most important restraints used in NMR structure calculations are the inter-proton distances derived from NOE-based experiments. In NOESY experiments, the NOE causes a change in the intensities of dipolar-coupled resonances. Thus, in a separate two-spin *IS* system, saturation of *S* resonance leads to the change in the intensity of *I* resonance, provided the inter-nuclear distance allows efficient dipolar cross-relaxation to occur. In such an ideal system, the maximum value of the NOE is independent of the inter-proton distance. However, the cross-relaxation rate constant σ_{IS}, that determines how quickly the NOE builds up, is inversely proportional to the sixth power of the inter-nuclear distance r_{IS}:

$$\sigma_{IS} \sim {r_{IS}}^{-6}$$

For practical reasons, the intensities of cross-peaks instead of cross-relaxation rate constants are used to derive inter-nuclear distances. This is possible because at the initial stage of the NOE build-up the size of the enhancement is controlled by the σ_{IS} (initial rate approximation) (Neuhaus and Williamson, 2000), and thus the NOE cross-peak intensity f_I after a certain time t is linearly correlated with σ_{IS} of the individual two-spin system:

$$f_I(t) \sim \sigma_{IS}$$

In principle, the inter-nuclear distance can be obtained by comparing σ_{IS} for the spin pair of interest (σ_{IS}^{int}) with σ_{IS} for a spin pair with known inter-nuclear distance (σ_{IS}^{ref})

(e.g. methylene or *ortho* aromatic protons) that acts as a reference (Neuhaus and Williamson, 2000):

$$\frac{\sigma_{IS}^{int}}{\sigma_{IS}^{ref}} = \frac{f_{int}(t)}{f_{ref}(t)} = \left(\frac{r_{IS}^{int}}{r_{IS}^{ref}}\right)^{-6} \implies r_{IS}^{int} = r_{IS}^{ref}\left(\frac{f_{int}(t)}{f_{ref}(t)}\right)^{-\frac{1}{6}}$$

However, the NOE-derived distances are intrinsically inaccurate (Neuhaus and Williamson, 2000) as:

1. The two spins are never in isolation and an indirect effect via spin diffusion can lead to significant calibration errors. Thus, the NOE observed on spin B when irradiating spin A can derive both from a direct A-B coupling and from indirect A-C and C-B couplings. If the A-C and B-C distances are significantly shorter than the A-B distance the contribution of the indirect pathway is not negligible.
2. The internal dynamics of macromolecules affects both cross-relaxation rate constants (σ_{IS}) and inter-nuclear distances (r_{IS}). The σ_{IS} is dependent on correlation time, that may be different for two spin pairs with identical r_{IS}. At the same time, the r_{IS} varies with time due to internal motions and this leads to averaging of σ_{IS} over all conformers for individual two-spin systems.
3. The reference distances are not accurate because they are also affected by internal mobility. Furthermore, they are significantly shorter than most of the unknown distances and therefore are not ideal for calibration.
4. The integration of cross-peak volumes or the measurement of their intensities can be compromised by spectral overlap or poor baseline.

Consequently, NOE-derived distances (r_{IS}) are considered as semi-quantitative and are typically specified by lower (r_{IS}^{lower}) and upper (r_{IS}^{upper}) bounds:

$$r_{IS}^{lower} \leq r_{IS} \leq r_{IS}^{upper}$$

In the structure calculations of FIR RRM1-RRM2 and FIR RRM1-RRM2 – FBP Nbox complex ~ 4500 NOE distance restraints were used, giving on the average 20 restraints per residue. The distance calibration was performed with automatic procedure implemented into structure calculation protocols. The process was successful for FIR RRM1-RRM2 alone, but failed for the complex, so that a converged family of

structures could have not been obtained. This is likely due to an intermediate regime of the exchange process between the free and bound states of the interacting partners, that causes broadening of the resonances. Consequently, the calibrated distances are weighted towards longer-than-real distances. To overcome this inaccuracy a semi-quantitative approach to distance calibration of inter-molecular restraints was applied.

2.1.3.2 Hydrogen bond distance restraints

H-bond distance restraints are powerful determinants of protein structure. The most common H-bond restraints used in protein structure determination are that between carbonyl (acceptor) and amide (donor) groups of the protein backbone. In α-helical regions, the H-bond occurs between an amide group of n residue and a carbonyl group of n-4 residue. In β-regions, the H-bond takes place between amide and carbonyl backbone groups in the adjacent β-strands within the β-sheet structure. NMR experiments that directly correlate H-bond donor and acceptor groups are of limited use in protein structure determination because of their low sensitivity due to the small coupling (1 – 2 Hz) (Grzesiek et al., 2001). A more widespread approach introduces H-bonds based on the initial rounds of structure calculations. This indirect identification is typically supported by information obtained from presence of donor and acceptor in regular secondary structure elements or, in the best case, from the protection of the amide donor protons against exchange with $^{2}H_2O$. In a $^{2}H_2O$ exchange experiment, the sample is lyophilised and then resuspended in $^{2}H_2O$, where the unprotected amide protons are rapidly exchanged with deuterons leading to a fast decrease of the intensity of the amide resonances in ^{1}H-^{15}N correlation spectra. The involvement of a ^{1}H in an H-bond slows the exchange rate very significantly:

$$-N-H\,|||||\,O=C< \;\rightleftarrows\; -N-H\quad O=C< \;\rightleftarrows\; -N-D\quad O=C<$$

It is normally assumed, that if a resonance is visible in the ^{1}H-^{15}N correlation spectra, the amide proton must be hydrogen bonded. In reality, protection factors of amide protons are not linked only to their involvement in H-bonds (Polshakov et al., 2006), but

measurement of ^2H_2O exchange rates, when feasible, represents a straightforward method to validate H-bonds identified during structure calculations procedure. A limitation of the method is that, in small and not very stable domains or regions of particular domains, domain 'opening' (unfolding and refolding) can be too fast to observe the effect even for the protected resonances.

In the structure calculations reported in this thesis H-bond distance restraints were applied, if an H-bond was observed in ≥ 50% of the structures obtained in the previous round of structure calculations and if the amide donor proton was protected against exchange with ^2H_2O. Interestingly, despite exchange experiments being performed at relatively high pH and temperature (pH 8.0, 37 °C), they showed the protection of ~ 80 HN groups in both FIR RRM1-RRM2 free and bound. However the indirect detection of H-bond acceptors did not allow me to identify unambiguously all of them. Most of the H-bonds identified using the method above are typical for α-helical and anti-parallel β-sheet structures. In the case of FBP Nbox peptide, the H-bonds restraints were introduced for the α-helical part of the peptide based on the NOE pattern and the values of dihedral angles. This is, because resonances of amide groups in all ^1H-^{15}N correlation spectra of the peptide (free and bound) at pH 8.0 and 37 °C are significantly broadened due to an intermediate regime of the exchange process of protons with water molecules and therefore ^2H_2O exchange experiments are not a feasible option.

2.1.3.3 Torsion angle restraints

The geometry of protein backbone can be described by three torsion angles: φ, ψ and ω that represent rotations around N-C^α, C^α-C′ and C′-N bonds, respectively. The peptide bond (C′-N) in proteins is planar due to its partial double-bond character and in most cases (> 99%) the corresponding ω angle is in *trans* conformation (~ 180°) (Sheik et al., 2003; Stewart et al., 1990). The values of the other two torsion angles (φ and ψ), although limited by steric hindrance, can assume different values (Sheik et al., 2003). These values define the geometry of the protein backbone, which is very different in the different secondary structure elements (Neuhaus and Williamson, 2000).

The values of φ and ψ torsion angles can be estimated experimentally from measurements of one-, two-, three- and four-bond scalar couplings (Edison et al., 1994; Vuister and Bax, 1994), with the most popular being three-bond $^{1}H^{N}$-$^{1}H^{\alpha}$ scalar coupling for estimation of φ angle (Vuister and Bax, 1993). However, a more common approach takes advantage of the empirical correlation between the local geometry of the protein backbone and the chemical shift values of both backbone $^{1}H^{\alpha}$, $^{13}C^{\alpha}$, $^{13}C'$ and ^{15}N resonances and the side-chain $^{13}C^{\beta}$ resonances. An important advantage of this approach is that the information required (the backbone and $^{13}C^{\beta}$ chemical shifts) are obtained during the assignment process and no additional NMR experiments are needed. Different protocols have been developed to derive torsion angle restraints from the chemical shifts. The one used in my studies is implemented in the TALOS (torsion angle likelihood obtained from shifts and sequence similarities) software (Cornilescu et al., 1999). The protocol utilises both sequence and chemical shift homology to predict the most likely backbone angles for a given residue. To improve the reliability of predictions a stretch of three consecutive residues (XYZ) is used to derive φ and ψ torsion angles of the centrally located residue (Y). The query triplet is compared with all triplets within the database consisting of 176 proteins with known high-resolution X-ray structure and known $^{13}C^{\alpha}$, $^{13}C^{\beta}$, $^{13}C'$, $^{1}H^{\alpha}$ and ^{15}N chemical shifts. The results are ranked according to the similarity in amino acid type and chemical shift values and then the prediction of the torsion angles is made based on the ten best matches. Importantly, if the torsion angles values are not consistent among the ten best scored triplets the prediction is discarded as unreliable. However good, TALOS is a database-based method and makes on the average ~ 3% of erroneous predictions and thus, care must be exerted when using the results obtained (Cornilescu et al., 1999).

TALOS protocol generated 236 angle predictions for FIR RRM1-RRM2 and 250 for FIR RRM1-RRM2 bound to the FBP Nbox peptide. Furthermore, 30 torsion angles were predicted for the peptide in the complex. The incorporation of these restraints into structure calculations improved the quality of obtained structures, especially at the initial stage of the calculations, when many NOE distance restraints were not unambiguously assigned.

2.1.4 Structure calculations – ARIA workflow

Several computational packages have been developed to assist structure determination from NMR data. The Ambiguous Restraints for Iterative Assignment (ARIA) has been designed to speed up the process of structure determination as well as to optimise structure quality. ARIA combines three powerful ideas: automated NOE assignment, usage of ambiguous NOE distance restraints and automated structure calculations (Güntert, 2003; Habeck et al., 2004; Linge et al., 2001; Nilges, 1996; Nilges et al., 1997; Nilges and O'Donoghue, 1998; Williamson and Craven, 2009). Thus, ARIA uses unassigned or partially assigned NOE peak lists with their intensities, together with protein sequence and chemical shift list, to perform NOE assignment, distance calibration, structure calculations, violation analysis, structure refinement and analysis of structure quality practically without any external intervention.

The principal aim of ARIA is to facilitate analysis of ambiguous NOEs and to incorporate this ambiguous data into automated structure calculations. The ambiguity of NOE cross-peaks is an inevitable consequence of increasing number of ^{1}H in larger proteins that entails an increased probability of having two or more nuclei with identical or very similar chemical shifts. Thus, a single cross-peak in the NOESY spectra, based only on chemical shifts, may be assigned to several nuclei. Among possible assignments there are real ambiguous (i.e. the peak represents contributions from at least two different inter-proton distances) as well as false assignments, that cannot be eliminated without any additional information. One solution to the false assignment problem is to calculate a preliminary structure only with unambiguous restraints and then try to resolve ambiguities based on the result of the structure calculations. Such a strategy is used during the non-automated iterative procedure of structure calculations. However, this procedure still discards correlations that derive from multiple contributions and that in some cases may be powerful restraints. ARIA takes advantage of these ambiguous correlations by treating each peak as a sum of all possible assignments and therefore as a sum of restraints, each weighted by the inverse sixth power of the corresponding inter-proton distance. Furthermore, an iterative protocol of ARIA for cross-peaks assignment, structure calculations and evaluation of assignment process allows elimination of incorrect assignments and thereby decreases the NOE ambiguity with each iteration step (Figure 7).

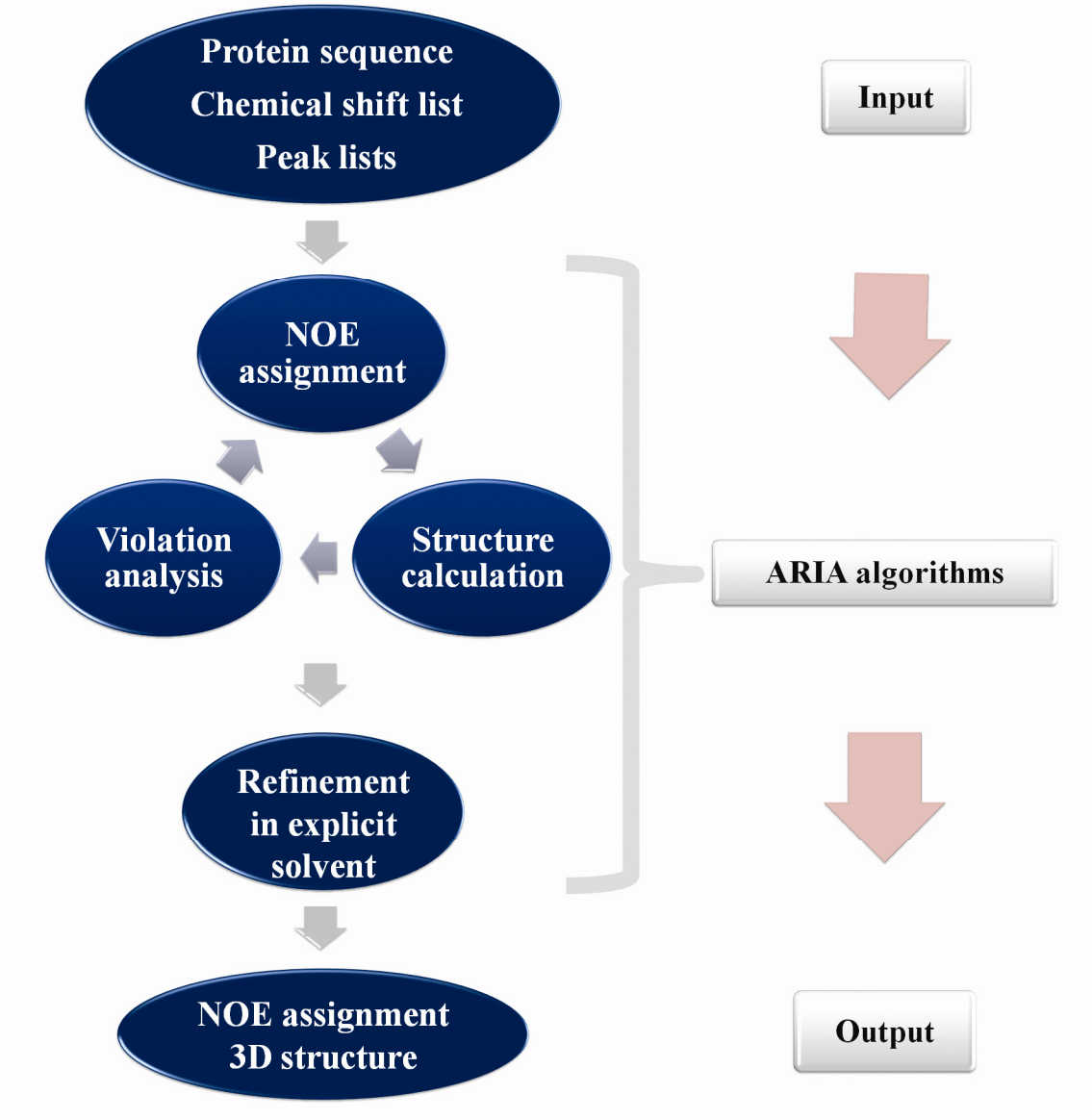

Figure 7 ARIA workflow.

ARIA uses as an input a protein sequence, a chemical shift assignment list and unassigned or partially assigned peak lists from NOESY spectra. In the first iteration (by convention called 'iteration 0'), the assignment of peaks is performed exclusively based on the chemical shift list. The assigned restraints are subsequently calibrated by ARIA. To perform this calibration ARIA does not use reference distances r_{IS}^{ref}, but derives a calibration factor C from the equation using the experimental data:

$$C = \sum_{NOEs} \frac{r_{IS}^{-6}}{V}$$

where:

C calibration factor

r_{IS} inter-proton distance

V observed NOE volume or intensity

The sum runs over all experimentally measured NOEs for which r_{IS} are shorter than a cut-off (6 Å). The ambiguous restraints are taken into account by replacing r_{IS} with an effective distance ('r_{IS}^{-6} summed distance') that contains contributions from all possible assignments. The obtained calibration factor is subsequently used to calculate the observed distance:

$$r_{IS}^{obs} = (CV)^{-1/6}$$

Importantly, the performance of this calibration method is rapid and very effective even when distances are derived from an extended template conformation (in iteration 0) of the polypeptide chain. For reasons discussed before (see section 2.1.3.1 NOE distance restraints) the calibrated restraints are intrinsically approximate and in ARIA the error Δ is estimated from an empirical equation:

$$\Delta = \varepsilon\left(r_{IS}^{obs}\right)^2 \qquad \text{(typically } \varepsilon = 0.125 \text{ Å}^{-1}\text{)}$$

Then the upper and lower bounds are defined by:

$$r_{IS}^{lower} = r_{IS}^{obs} - \Delta$$
$$r_{IS}^{upper} = r_{IS}^{obs} + \Delta$$

The restraints obtained are subsequently used to calculate structures. The aim of structure calculation protocols is to find a global minimum region of the total energy function E_{hybrid} of the system. This function, also called force field, combines the NMR experimental data and the information on covalent geometry (bonds, angles, planarity, chirality) and the non-bonded contacts (van der Waals force). The individual terms are represented as energy contributions and are accordingly weighted:

$$E_{hybrid} = \sum_i w_i E_i$$

where:

E_i individual energy contributions

w_i weight factors for each energy term

In ARIA, the global minimum of total energy function E_{hybrid} is calculated with the Crystallography and NMR System (CNS) (Brunger et al., 1998), that relies on molecular dynamics-based simulated annealing protocol. The molecular dynamics protocol calculates the trajectories of all atoms within the molecule using the equations of Newtonian mechanics and treating the molecule essentially as a sophisticated mechanical model. Initially, every atom obtains some kinetic energy in the form of an initial velocity in a random direction – this is done based on a Maxwellian distribution corresponding to a notional temperature for the molecule as a whole. The movement of every atom is affected by different forces that are represented in the force field. Considering these complex relations, the relative positions of atoms and their velocities are recalculated at very short time intervals (i.e. short relative to the highest frequency of motion in the system) simulating the movements of the atoms over time. The simulations are performed in the torsion angle space that significantly reduces the freedom of atoms, as the only degree of freedom are the torsion angles. Furthermore, the protocol includes a high-temperature search phase followed by a slow cooling phase, which slowly reduces the kinetic energy of the atoms. Such an approach minimise the problem of trapping in the local minimum of the E_{hybrid} function. Importantly, the ARIA protocols change some of the parameters during calculations in

a way that has been optimised to maximise the possibility of reaching a global minimum of the force field.

Following structure calculations, violation analysis is performed to eliminate noise-derived and erroneously assigned peaks. Afterwards the next iteration starts, but this time assignment of ambiguous restraints is based on the best fraction of calculated structures in the preceding iteration. The whole procedure continues for several iterations (typically nine) during which parameters for assignment of ambiguous restraints and violation analysis are being tighten up to improve the quality of calculated structures.

In addition to NOE-derived distance restraints ARIA can use complementary experimental information such as hydrogen bonds, torsion angles, residual dipolar couplings and disulfide bridges.

Upon completion of the last iteration (typically iteration 8) the final structures are a subject to a short molecular dynamics refinement in explicit solvent with a full force field, including electrostatic and Lennard-Jones potentials. This step improves the quality of obtained structures as they often show artefacts due to simplified treatment of non-bonded forces and missing solvent contacts during structure calculations.

Finally, ARIA generates many output files that report on performance of assignment and structure calculations procedures as well as on quality of the final conformers.

2.1.5 Strategy for the structure determination of protein–protein complexes by NMR

NMR spectroscopy is a powerful tool in the study of macromolecular interactions in solution and has been proved to be particularly useful for the structural investigation of transient complexes that are usually not amenable by X-ray techniques (Nietlispach et al., 2004). NMR-driven structure determination is limited by the molecular weight of the studied molecule/complex. Typically, complexes with molecular weight > 50 kDa have large rotational correlation time (τ_c) and thereby,

the sensitivity of most multi-dimensional NMR experiments is low. Furthermore, the large number of ^{1}Hs in high molecular weight complexes results in the overlap of their resonances and hampers the assignment process. A common strategy to overcome this problem is to identify which parts/domains of the binding partners are involved in the interaction and to reduce the system under study to these regions (Nietlispach et al., 2004).

Ideally, two samples – each containing one of the components labelled and the other unlabelled – are used to complete the assignment of the complex. Two separate sets of experiments are recorded for the two samples. The comparison of these experiments allows one to identify intra- and inter-molecular restraints and the assignment of inter-molecular restraints is then validated using filtered experiments (Nietlispach et al., 2004). In some cases, the small size of one of the two components allows for the structure determination of the complex without that component to be labelled (Nietlispach et al., 2004).

In practice the strength of the inter-molecular interaction often defines the strategy used for the structure determination of the complex (Nietlispach et al., 2004). Complexes with dissociation constants in the micromolar range, such as the one between FIR RRM1-RRM2 and FBP Nbox usually have resonances in the moderately fast exchange regime on a chemical shift time scale (depending on the changes in chemical shifts between the free and bound states). The ratio of the two components during experiments on the complex is chosen so that the observed component is nearly-fully saturated, which typically requires an excess of the other component. Saturation of the observed component serves three purposes: firstly, it minimises broadening due to exchange between the free and bound conformation; secondly, it allows maximisation of the signal coming from the bound component and thirdly, it helps to maintain consistency in the chemical shift of references in the bound form. Interestingly, relatively weak binding can be of help during the assignment of the bound resonances because the shift of the well dispersed ^{1}H, ^{13}C and ^{15}N resonances during titration with the other component can be used to transfer assignments from the free to the bound form, without having to undertake a full assignment procedure. It is also worth mentioning that complexes in very and moderately fast exchange regime are often difficult to crystallise. This makes NMR spectroscopy techniques especially useful in studies of such transient complexes.

Preliminary NMR experiments revealed that the FIR–FBP–FUSE system is indeed in the fast-to-intermediate exchange regime and that the affinity of FIR for ssFUSE and FBP Nbox is in the micromolar range. Furthermore, the previously performed studies (discussed in section 1.8 Interactions between FUSE, FBP and FIR) and my NMR experiments showed that it is possible to reduce the system under study to a size amenable for NMR for high resolution studies. Thus, NMR is the technique of choice to study the structure of these transient complexes.

2.2 Circular Dichroism

Circular dichroism (CD) is a spectroscopic technique that allows one to obtain low-resolution structural information about proteins and nucleic acids from small amounts of material (Martin and Schilstra, 2008). CD spectroscopy measures the difference in absorbance for left- and right-handed circularly polarised light. The CD signal derives from chromophores, the same as those that give rise to absorbance and fluorescence phenomena, but to observe the CD signal these chromophores have to be inherently chiral or to be located in a chiral environment (Martin and Schilstra, 2008). In proteins, chromophores are generally achiral and the observed signal is generated by the chiral environment of the protein. In the near-UV CD spectra of proteins the signal originates from tryptophan, tyrosine, phenylalanine and cystine and reflects the tertiary, and sometimes quaternary, structure of the protein. In the case of the far-UV CD spectrum of proteins the absorbance phenomenon comes mainly from $n \rightarrow \pi^*$ (at ~ 220 nm) and $\pi \rightarrow \pi^*$ (at ~ 190 nm) transitions within the peptide bond and reflects the secondary structure of the protein (Martin and Schilstra, 2008).

The main advantages of CD are high sensitivity to changes in conformation, the possibility to study a broad range of solvent conditions and requirement for small amounts of the material (Martin and Schilstra, 2008). CD has been extensively used to 1) estimate the secondary structure content of the proteins, 2) detect conformational changes within the macromolecules, 3) study folding, unfolding and stability of the macromolecules and 4) study inter-molecular interactions (Martin and Schilstra,

2008). In this thesis, CD was employed to evaluate the fold (from the far-UV CD spectrum) and stability (from the thermal denaturation experiment) of the FIR RRM1-RRM2, and also to estimate the secondary structure content of the FBPs Nbox peptides.

2.3 BioLayer Interferometry

BLI is a label-free surface-based optical analytical technique to study inter-molecular interactions. The detection is based on the interference of electromagnetic waves (from the visible light spectrum) that are partially reflected from an internal reference layer and from a layer of immobilised molecules (biolayer). The reflected waves encounter constructive or destructive interference, depending on the wavelength, and create a characteristic intensity pattern across the white light spectrum (Figure 8). When the thickness of the biolayer increases due to inter-molecular interactions on the biosensor surface a shift in the interference pattern is observed.

BLI has been implemented in the Octet platform (ForteBio) that allows eight parallel binding assays being performed in a relatively short time (a few minutes to a few hours, depending on the set-up). The researcher's intervention is limited to preparing solutions with the desired compositions in a 96-well plate, choosing the parameters of the assay and analysing the data. A "dip and read" technology eliminates problems of microfluidics becoming clogged and significantly reduces the necessary maintenance. Furthermore, disposable sensors eliminate the need for surface regeneration, significantly increasing throughput and allowing a quick test and/or optimisation of the assay conditions.

In the course of a surface-based experiment (e.g. in the Octet platform) one of the interaction partners is immobilised on the surface of a tip (e.g. by streptavidin-biotin binding) which is then exposed to a solution containing the other interaction partner. Binding changes the thickness of the biolayer generating the detectable signal. The main pitfall of surface-based optical analytical techniques is the possibility of non-specific binding of the non-immobilised protein to the surface. This can be tested in a control experiment where the first component is not immobilised on the surface and the tip is exposed to the second component. Preliminary BLI experiments on FUSE–FBP–FIR

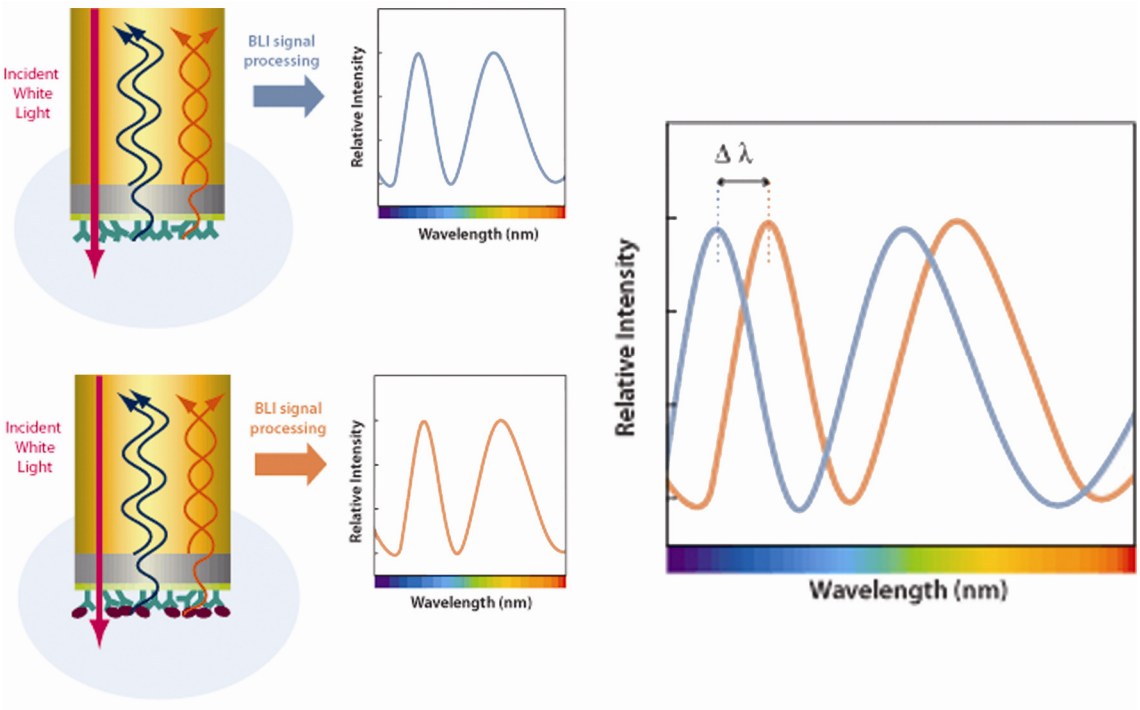

Figure 8 Physical background of BLI method (figure taken from http://www.fortebio.com/bli_technology.html). The detected signal originates from an interference between electromagnetic waves reflected from an internal reference surface and an immobilised-molecules surface. Upon inter-molecular interaction a distance between two reflecting surfaces increases causing a shift in an interference pattern that is a detected signal.

system, which were performed in the buffer used in the NMR titrations (10 mM Tris-HCl pH 7.4, 50 mM NaCl, 2 mM Tris(2-carboxyethyl) phosphine (TCEP)), showed that the FBP constructs bind non-specifically to the sensor surface in the absence of immobilised ssDNA. Thus, the experimental conditions were optimised and it was found that increasing the salt concentration (up to 150 mM) and including bovine serum albumin (BSA) (0.5 mg/ml) in the buffer eliminates these non-specific interactions.

2.4 Experimental details

2.4.1 Cloning

The cDNA sequences encoding FIR RRM1, FIR RRM2 and FIR RRM1-RRM2 (NP_055096) were PCR-amplified from the pCMV-rSiahBP plasmid (a gift of Dr. S. Kindler) using primers that introduced 5' NcoI and 3' HindIII restriction sites (Table 1). The original plasmid carries a sequence of rat homologue of FIR, rSiahBP (AF165892). However, the protein is very conserved and no differences exist in the region of interest (RRM1-RRM2) at the protein level (Appendix I). Thus, in this thesis, the numbering of protein residues is after the human FIR (NP_055096). The PCR products were cloned between NcoI and HindIII sites of the pETM-30 expression vector (EMBL-Heidelberg, Protein Expression Facility) generating plasmids, which encode four variants of FIR RRM1, FIR RRM2 and FIR RRM1-RRM2 (Table 2, see also Figure 10) with an N-terminal tobacco etch virus (TEV) protease-cleavable hexa-histidine tag (HisTag) glutathione-S-transferase (GST) fusion.

The cloning of FBP constructs described below has been performed by David Hollingworth at the National Institute for Medical Research, London, UK.

The cDNA sequence encoding full length FBP sequence was amplified and cloned from a cDNA library (MegaMan human transcriptome library, Stratagene) following the manufacturer's instructions and using FBP-F and FBP-R primers (Table 1). FBP Nbox-KH1-KH4 (amino acids 27-455, NP_003893) and FBP KH1-KH4 (amino acids 85-455, NP_003893) were PCR-amplified from the full-length clone using

Table 1 Primers used to clone FIR and FBP constructs.

Name	Sequence	Forward /Reverse	Restriction site	Nucleotides[a]
FIR				
F-RRM1-L	5'-AGA TCC ATG GCT CAG CGG CAG CGG GCA C-3'	Forward	NcoI	319-337
F-RRM1-S	5'-AGA TCC ATG GCT ATC ATG TGC CGG GTG TAT G-3'	Forward	NcoI	340-361
F-RRM2-L	5'-AGA TCC ATG GCT AGC AAC ATC GGA CAA GCC CA-3'	Forward	NcoI	578-599
F-RRM2-S	5'-AGA TCC ATG GCT TTC AAC CGA ATA TAC GTG G-3'	Forward	NcoI	633-655
R-RRM1-L	5'-A GAT AAG CTT *TCA* AGC CCT AGC CTC CTC AGC CA-3'	Reverse	HindIII	617-637
R-RRM1-S	5'-A GAT AAG CTT *TCA* AGG TCT TCC CAC CTT GAT GTT C-3'	Reverse	HindIII	558-579
R-RRM2-L	5'-A GAT AAG CTT *TCA* CGT GGC AGG TGT TAG CAG GG-3'	Reverse	HindIII	887-906
R-RRM2-S	5'-A GAT AAG CTT *TCA* TGT GAC GGC CTT GCC CAC CC-3'	Reverse	HindIII	857-876
FBP				
FBP-F	5′-CTT GTC TAG ATG GCA GAC TAT TCA ACA GTG CCT-3′	Forward	XbaI	90-113
FBP-R	5′-C TTG CTC GAG *TTA* TTG GCC CTG AGG TGC TGG AG-3′	Reverse	XhoI	2002-2021
FBP-F-Nbox	5′-CTT GCC ATG GGA GTT AAC GAC GCT TTC AAA GA-3′	Forward	NcoI	165-187

Name	Sequence	Forward /Reverse	Restriction site	Nucleotides[a]
FBP				
FBP-F-KH1	5′-CTT GCC ATG GGA TTT GGA ACA CAG TTA CCA CCG ATG CA-3′	Forward	NcoI	342-367
FBP-R-KH4	5′-CT TGA AGC *TTA* CCC TAA AGG ATT TAC TGG GCC A-3′	Reverse	HindIII	1433-1454
Nbox-F	5′-GCA TCC ATG GGC TAT GTG AAC GAT GCG TTT AAA GAT GCG CTG CAG CGC GCG CGC CAG ATT GCG GCG AAA ATT GGC GGC GAT GCG GGC ACC AGC *TAA* GCT TAT CG-3′	Forward	NcoI	-[b]
Nbox-R	5′-CG ATA AGC *TTA* GCT GGT GCC CGC ATC GCC GCC AAT TTT CGC CGC AAT CTG GCG CGC GCG CTG CAG CGC ATC TTT AAA CGC ATC GTT CAC ATA GCC CAT GGA TGC-3′	Reverse	HindIII	-[b]

[a] corresponding nucleotides in the cDNA sequence: AF165892 for FIR and NM_003902 for FBP

[b] these primers were not used in PCR, but paired together and cloned into the expression vector

Restriction sites are underlined and STOP codons are in italic

Table 2 Summary of the FIR and FBP constructs.

Construct name	Protein	Residues cloned[a]	Forward primer	Reverse primer
FIR RRM1-RRM2[b]	FIR	103-297	F-RRM1-L	R-RRM2-L
FIR RRM1-RRM2 (103-287)	FIR	103-287	F-RRM1-L	R-RRM2-S
FIR RRM1-RRM2 (110-297)	FIR	110-297	F-RRM1-S	R-RRM2-L
FIR RRM1-RRM2 (110-287)	FIR	110-287	F-RRM1-S	R-RRM2-S
FIR RRM1 (103-207)	FIR	103-207	F-RRM1-L	R-RRM1-L
FIR RRM1 (103-188)	FIR	103-188	F-RRM1-L	R-RRM1-S
FIR RRM1 (110-207)	FIR	110-207	F-RRM1-S	R-RRM1-L
FIR RRM1 (110-188)	FIR	110-188	F-RRM1-S	R-RRM1-S
FIR RRM2 (189-297)	FIR	189-297	F-RRM2-L	R-RRM2-L
FIR RRM2 (189-287)	FIR	189-287	F-RRM2-L	R-RRM2-S
FIR RRM2 (208-297)	FIR	208-297	F-RRM2-S	R-RRM2-L
FIR RRM2 (208-287)	FIR	208-287	F-RRM2-S	R-RRM2-S
FBP Nbox-KH1-KH4	FBP	27-455	FBP-F-Nbox	FBP-R-KH4
FBP KH1-KH4	FBP	85-455	FBP-F-KH1	FBP-R-KH4
FBP Nbox	FBP	27-52	Nbox-F	Nbox-R

[a] numbering according to NP_055096 for FIR and NP_003893 for FBP

[b] main construct of FIR used

the primers that introduced 5' NcoI and 3' HindIII restriction sites (Tables 1 and 2). The PCR products were further processed as described above for FIR.

The FBP Nbox (amino acids 27-52, NP_003893) expression plasmid was generated by pairing two complementary oligonucleotides (Tables 1 and 2) and their subsequent cloning into pETM-30 vector as described above. The sequence of oligonucleotides was optimised for expression in *E. coli* cells and introduced N-terminal Y residue to facilitate quantification of the recombinant peptide.

The nucleotide sequence of the expression plasmids was confirmed by DNA sequencing (Lark Technologies).

2.4.2 Protein expression

Expression of the FBP Nbox-KH1-KH4 and FBP KH1-KH4 proteins has been performed by David Hollingworth at the National Institute for Medical Research, London, UK.

Unlabelled recombinant FBP and FIR proteins were expressed in *E. coli* BL21 (DE3) cells (Invitrogen) growing in Luria-Bertani broth (LB medium) with kanamycin (50 μg/ml) (Sigma). 25 μl of the cells were transformed using a standard heat shock protocol with 0.5 μl of the expression plasmid (~ 20 ng/ml). 100-300 μl of transformed cells were used to inoculate 100 ml starting cultures in LB media. The cells were grown at 37 °C until the optical density at 600 nm (OD_{600}) reached a value of 1. 50 ml of the starting cultures were used to inoculate 1.25 l of LB media and the cells were grown at 37 °C until the OD_{600} = 0.6. Protein expression was induced with isopropyl β-D-1-thiogalactopyranoside (IPTG) (0.5 mM final concentration) (Sigma). The cells were cultured for further four hours, harvested by centrifugation (Beckman J2-21 centrifuge, rotor JA-10, 7000 rpm, 10 min., 4 °C) and stored at -80 °C. Labelled (^{15}N or $^{15}N^{13}C$) FIR RRM1-RRM2, (^{15}N) FIR RRM2 and (^{15}N or $^{15}N^{13}C$) FBP Nbox were obtained by expressing the constructs in cells growing in M9 minimal media containing $^{13}C_6$-D-glucose (Cambridge Isotope Laboratories) as the only source of carbon and/or $^{15}NH_4Cl$ or $(^{15}NH_4)_2SO_4$ (Cambridge Isotope Laboratories) as the only source of

nitrogen. $^{15}N^{13}C^{2}H$-labelled FIR RRM1-RRM2 was obtained by growing the cells in a $^{2}H_2O$ (GOSS Scientific Instruments Ltd.) solution of the media above.

2.4.3 Protein purification

Purification of FBP Nbox peptide on a Zorbax 300 SB-C_{18} column and purification of FBP protein constructs have been performed by David Hollingworth at the National Institute for Medical Research, London, UK.

Frozen cells were resuspended in Equilibration buffer (10 mM Tris-HCl pH 8.0, 10 mM imidazole, 200 mM NaCl, 2 mM β-mercaptoethanol) (10 ml per 1 l of cell culture) with 0.5% (v/v) Triton X-100, DNase I (Roche) and lysozyme (Sigma), sonicated on ice (Branson Sonifier 250, power output 90 W, 90% duty cycle, 3 x 40 s) and centrifuged (Beckman J2-21 centrifuge, rotor JA-20, 18000 rpm, 40 min., 4 °C).

The overexpressed recombinant HisTag-GST-FIR RRM1-RRM2 and HisTag-GST-FIR RRM2 proteins were purified from the soluble fraction using immobilised metal ion affinity chromatography (IMAC) columns (Qiagen) according to the manufacturer's instructions. Briefly, the soluble fraction was applied onto the IMAC column (5 ml volume, gravity flow) equilibrated with 50 ml of Equilibration buffer. Next, the column was washed (50 ml of Washing buffer – 10 mM Tris-HCl pH 8.0, 10 mM imidazole, 1 M NaCl, 2 mM β-mercaptoethanol) and the protein was eluted with 25 ml steps of 30, 100 and 300 mM imidazole in 10 mM Tris-HCl pH 8.0, 1 M NaCl, 2 mM β-mercaptoethanol (Elution buffers 30, 100 and 300). The majority of the fusion proteins was found in 100 mM imidazole fractions. The HisTag-GST fusion was removed by overnight TEV protease digestion (2.5 μM enzyme) at 4 °C during dialysis against Equilibration buffer, followed by a second IMAC step. Briefly, the digested protein was applied onto the equilibrated (50 ml of Equilibration buffer) IMAC column. Once the solution passed through, additional 5 ml of Equilibration buffer was added to wash out the recombinant protein not bound to the column and then the flow-through fraction was further purified passing it through the second IMAC column. The FIR RRM1-RRM2 and FIR RRM2 solutions were concentrated (final volume < 5 ml) in an Amicon stirred cell (10 kDa cut-off membrane, Millipore) and purified by size-

exclusion chromatography (SEC) (HiLoad Superdex 75 16/60 column, Pharmacia) in 10 mM Tris-HCl pH 8.0, 200 mM NaCl, 2 mM β-mercaptoethanol. The SEC step was omitted in one of the preparations (see Results). The pure proteins (> 95 % as assessed by SDS-PAGE and Coomassie staining) were dialysed against 10 mM Tris-HCl pH 8.0 (or pH 7.4), 50 mM NaCl, 100 μM TCEP, concentrated in an Amicon stirred cell (10 kDa cut-off membrane, Millipore) to 0.4-0.8 mM and stored in 10 mM Tris-HCl pH 8.0 (or pH 7.4), 50 mM NaCl, 2 mM TCEP, 0.05% (w/v) NaN_3 in the presence of protease inhibitors (Roche) at -80 °C.

Unlabelled FBPs Nbox peptides (Table 3), with and without an N-terminal Y residue, were chemically synthesised (Peptide Synthesis Facility, University of Bristol and in-house, respectively). The peptides were resuspended in 10 mM Tris-HCl pH 7.4 at concentration ~ 2 mM and the pH was adjusted to 7.4 with NaOH. The addition of Y residue does not affect K_ds of FBPs Nbox – FIR RRM1-RRM2 complexes as assessed by NMR titrations (14 ± 8 μM *vs.* 8 ± 5 μM for FBP Nbox and 6 ± 6 μM *vs.* 13 ± 6 μM for FBP2 Nbox). Thus, for simplicity in the following results the peptides are referred as FBP Nbox and FBP2 Nbox for both the peptides with and without Y residue.

Labelled (^{15}N or $^{15}N^{13}C$) FBP Nbox peptide (with N-terminal Y) was purified by IMAC as described above (the peptide was found in 100 and 300 mM elution fractions), dialysed against 10 mM Tris-HCl pH 8.0, 10 mM NaCl and applied to a Zorbax 300 SB-C_{18} column (250 x 9.4 mm I.D.; 5-μm particle size, 300-Å pore size; Agilent Technologies) at a flow rate of 2 ml/min. and temperature of 40 °C. Initial run conditions were 2% eluent B, followed by a step to 20%. The desired peptide was then eluted with a linear A-B gradient from 20% B to 55% B over 45 minutes (0.35% acetonitrile/min.), where eluent A was 0.1% aqueous trifluoroacetic acid (TFA) in water and B was 0.1% TFA in 90% acetonitrile. Desired peak was collected, lyophilised, resuspended in water and re-lyophilised. The peptide was stored in 10 mM Tris-HCl pH 5.0 (or 8.0), 0.05% (w/v) NaN_3 in the presence of protease inhibitors (Roche) at -80 °C.

FBP Nbox-KH1-KH4 and FBP KH1-KH4 were purified by IMAC as described above. The IMAC-purified proteins, still containing a high nucleic acid content, were concentrated, buffer exchanged into 10 mM Tris-HCl pH 7.4, 50 mM NaCl, 2 mM β-mercaptoethanol, and applied to a MonoQ 5/50GL anion exchange column

Table 3 Chemically synthesised peptides recapitulating the Nbox peptides of FBP proteins.

Peptide	**Amino acid sequence**[a]	**Residue number in the protein**
FBP Nbox	(Y)VNDAFKDALQRARQIAAKIGGDAGTS	27-52
FBP2 Nbox	(Y)RKDAFADAVQRARQIAAKIGGDAATT	70-95
FBP3 Nbox	KAEGFVDALHRVRQIAAKIDSIPHLN	15-40

[a] the N-terminal Y residue added to the FBP and FBP2 Nboxes does not affect their affinity for FIR RRM1-RRM2 as assessed by NMR titrations

(GE Healthcare) equilibrated in the same buffer. The flow-through, which contained protein only, was collected, concentrated and buffer exchanged into 10 mM Tris-HCl pH 7.4, 20 mM NaCl, 2 mM β-mercaptoethanol. This protein was applied to a 1 ml HiTrap Heparin column (GE Healthcare). The protein product was eluted using a linear gradient of 20 mM NaCl to 100 mM NaCl in the same buffer, over 30 minutes at a flow rate of 1 ml/min. The proteins were stored in 10 mM Tris-HCl pH 7.4, 50 mM NaCl, 2 mM TCEP, 0.05% (w/v) NaN_3 in the presence of protease inhibitors (Roche) at -80 °C.

Proteins and peptides concentrations were determined from the absorbance at 280 nm and their molecular weights and purity were confirmed by electrospray mass spectrometry (in-house). Concentrations of peptides without Y were determined from the absorbance at 258 nm.

2.4.4 ssDNA oligonucleotides

All ssDNA oligonucleotides (Table 4) and Scaffold-Independent Analysis (SIA) oligonucleotides pools were chemically synthesised (Sigma and Integrated DNA Technologies).

2.4.5 Circular dichroism

All CD experiments were performed on a Jasco J-715 spectropolarimeter (Jasco) equipped with a PTC-348 Peltier temperature-control system. The samples of FIR RRM1-RRM2 were prepared in 10 mM Tris-HCl pH 7.4, 50 mM NaCl and 2 mM TCEP at 0.14 mg/ml (a far-UV spectrum) or 0.27 mg/ml (thermal unfolding) concentrations. A far-UV spectrum of FIR RRM1-RRM2 was recorded at 25 °C. For thermal unfolding, a solution of RRM1-RRM2 was gradually heated (from 2 °C to 65 °C, 2 °C/min., step = 0.1 °C) and the unfolding was monitored at 220 nm. Far-UV CD spectra of 50 μM FBPs Nbox peptides in 10 mM Tris-HCl pH 7.4, 20 mM NaCl

Table 4 Chemically synthesised ssDNA oligonucleotides used in the binding studies.

Oligonucleotide name	Sequence
ssFUSE87[a]	5′-TGCT CCCTG AAATG ATCTA TATTT AATAT *ATAAT GTATA TTCCC TCGGG ATTTT TTATT TTGTG TTATT* CCACG GCATG AAAAA CAA-3′
ssFUSE40	5′-ATAAT GTATA TTCCC TCGGG ATTTT TTATT TTGTG TTATT-3′
ssFUSE6-40	5′-GTATA TTCCC TCGGG ATTTT TTATT TTGTG TTATT-3′
ssFUSE11-40	5′-TTCCC TCGGG ATTTT TTATT TTGTG TTATT-3′
ssFUSE16-40	5′-TCGGG ATTTT TTATT TTGTG TTATT-3′
ssFUSE1-35	5′-ATAAT GTATA TTCCC TCGGG ATTTT TTATT TTGTG-3′
ssFUSE1-30	5′-ATAAT GTATA TTCCC TCGGG ATTTT TTATT-3′
ssFUSE1-25	5′-ATAAT GTATA TTCCC TCGGG ATTTT-3′
ssFUSE29	5'-GTATA TTCCC TCGGG ATTTT TTATT TTGT-3′
TGTGT	5′-TGTGT-3′
TTTTT	5′-TTTTT-3′
TATAT	5′-TATAT-3′
AAAAA	5′-AAAAA-3′

[a] ssFUSE87 corresponds to nucleotides 751-837 of non-coding strand of the *c-myc* gene (X00364); nucleotides in italic recapitulate the ssFUSE40 within ssFUSE87

were recorded at 5 °C. The CD signal is reported as the mean residue CD extinction coefficient ($\Delta\varepsilon_{MRW}$). The secondary structure content of the different peptides was calculated using the SELCON, CDSSTR and CONTIN/LL programs, that are available online (http://lamar.colostate.edu/~sreeram/CDPro/).

2.4.6 NMR spectroscopy

Labelled (^{15}N, $^{15}N^{13}C$ or $^{15}N^{13}C^{2}H$) samples of FIR RRM1-RRM2 were prepared in 90% H_2O/10% $^{2}H_2O$ solutions of 10 mM Tris-HCl pH 8.0, 50 mM NaCl, 2 mM TCEP at concentrations in the range 0.25-0.35 mM. Samples of $^{15}N^{13}C$-labelled FBP Nbox were prepared in 90% H_2O/10% $^{2}H_2O$ solutions of 10 mM Tris-HCl pH 5.0 at concentrations in the range 0.35-0.4 mM. Samples of FIR RRM1-RRM2 – FBP Nbox complex (labelled : unlabelled ratio = 1 : 2 – 1 : 3 where one of the complex component is ^{15}N- or $^{15}N^{13}C$-labelled) were prepared in 90% H_2O/10% $^{2}H_2O$ or 100% $^{2}H_2O$ solutions of 10 mM Tris-HCl pH 8.0, 50 mM NaCl, 2 mM TCEP. The concentration of the labelled component was in the range 0.2-0.6 mM. Sample of ^{15}N-labelled FIR RRM1-RRM2 – unlabelled ssFUSE29 at 0.4 mM : 0.2 mM ratio was prepared in 90% H_2O/10% $^{2}H_2O$ solution of 10 mM Tris-HCl pH 7.4, 50 mM NaCl, 2 mM TCEP.

NMR spectra were recorded on ^{15}N-, $^{15}N^{13}C$- or $^{15}N^{13}C^{2}H$-labelled samples at 37 °C or 45 °C on Varian Inova and Bruker Avance spectrometers equipped with cryoprobes and operating at 600, 700 and 800 MHz ^{1}H frequencies. The spectra were processed with the NMRPipe package (Delaglio et al., 1995) and analysed with Sparky (Goddard and Kneller, 2004).

$^{1}H^{N}$, ^{15}N, $^{13}C^{\alpha}$, $^{13}C^{\beta}$ and $^{13}C'$ assignments were obtained from standard HNCACB, CBCA(CO)NH, HN(CO)CACB, HNCA, HN(CO)CA and HNCO backbone experiments (Bax and Grzesiek, 1993; Yamazaki et al., 1994). Side-chain resonance assignment was achieved from HCCH TOCSY (mixing time = 12 ms) (Kay et al., 1993), ^{15}N TOCSY (mixing time = 60 ms) (Marion et al., 1989) and ^{15}N and ^{13}C NOESY HSQC experiments (mixing time = 100 ms) (Fesik and Zuiderweg, 1988) optimised to observe aliphatic or aromatic resonances. The distance restraints for structure calculations were obtained from above mentioned NOESY HSQC experiments

and ^{13}C-filtered ^{13}C-edited NOESY HSQC (mixing time = 100 ms) (Zwahlen et al., 1997).

$^{1}H^{N}$ exchange protection data were obtained from sofast ^{15}N HMQC (Schanda et al., 2005) recorded immediately after resuspension of ^{15}N-labelled FIR RRM1-RRM2 or ^{15}N-labelled FIR RRM1-RRM2 – unlabelled FBP Nbox complex in 100% $^{2}H_2O$ solution of 10 mM Tris-HCl pH 8.0, 50 mM NaCl, 2 mM TCEP.

T1, T2 and hetNOE parameters were obtained from standard experiments (Kay et al., 1989) recorded on ^{15}N-labelled samples at 37 °C and 600 MHz ^{1}H frequency and analysed by using NMRPipe routines (Delaglio et al., 1995). T1/T2 ratio of residues in well-defined secondary structure elements were used to estimate the rotational correlation times of proteins and complexes with the program ModelFree (http://biochemistry.hs.columbia.edu/labs/palmer/software/modelfree.html).

Samples, temperature, solvent and ^{1}H frequency of spectrometers used for individual 3D and relaxation experiments can be found in Appendix II.

2.4.7 Structure calculations

Distance and angle restraints were used to perform structure calculations with ARIA 1.2 (Linge et al., 2001). Experimental distance restraints were achieved by integrating Sparky assigned NOE peaks with XEASY (Bartels et al., 1995). Dihedral restraints (φ and ψ) were obtained from the chemical-shift-based TALOS database (Cornilescu et al., 1999). For the protein, H-bond restraints were added in subsequent calculations if an exchange-protected $^{1}H^{N}$ was H-bonded in at least 50% of the preliminary structures generated in a given run. For the peptide, H-bond restraints were added for the α-helical region of the peptide (A30-K44).

In iterations 0-7 of ARIA calculations fifty randomised conformers were subject to simulated annealing with a standard CNS protocol (Brunger et al., 1998). The ten lowest-global-energy structures were used for assignment at the next iteration. In iteration 8 the number of generated structures was increased to 200. Finally, the 20 lowest-energy obtained structures were water refined (Linge et al., 2003). For the structure calculations of the complex, the automatic distance calibration was

turned off and the inter-molecular distances were calibrated semi-quantitatively (short = 2.8 Å, medium = 3.5 Å, long = 5.0 Å and very long = 6.0 Å), while the upper bounds for all intra-molecular restraints were set up to 6 Å. The quality of each generated family was evaluated with PROCHECK_NMR (Laskowski et al., 1996). The structures were displayed and analysed with MOLMOL (Koradi et al., 1996), PYMOL (www.pymol.org) and InsightII (Accelrys). Families of FIR RRM1-RRM2 and FIR RRM1-RRM2 – FBP Nbox complex were deposited in the PDB with the accession codes 2KXF and 2KXH, respectively.

2.4.8 Binding assays

2.4.8.1 NMR

25-75 μM ^{15}N-labelled samples of FIR RRM1-RRM2 in 10 mM Tris-HCl pH 7.4, 50 mM NaCl, 2 mM TCEP were titrated with unlabelled FBP Nbox or ssDNA oligonucleotides (29mer or 5mers). Titrations of ^{15}N- and ^{15}N^{13}C-labelled FBP Nbox with unlabelled FIR RRM1-RRM2 were performed in 10 mM Tris-HCl pH 5.0, 6.5 or 8.0, 50 mM NaCl, 2 mM TCEP at 50-70 μM peptide concentrations. For the tri-molecular titrations, saturated complexes of 25-75 μM ^{15}N-labelled FIR RRM1-RRM2 with ssFUSE29 (1 : 1.5 ratio), 5'-TGTGT-3' oligonucleotide (1 : 8 ratio) or FBP Nbox (1 : 6 ratio) in 10 mM Tris-HCl pH 7.4, 50 mM NaCl, 2 mM TCEP were titrated with FBP Nbox or 5'-TGTGT-3' oligonucleotide. ^{15}N sofast HMQC (Schanda et al., 2005), or ^{13}C HSQC (Palmer et al., 1991) spectra were recorded at each titration point at 37 °C or 45 °C on Varian Inova and Bruker Avance spectrometers equipped with cryoprobes and operating at 600, 700 and 800 MHz ^{1}H frequencies. The set-ups for individual titration experiments can be found in Appendix III. To obtain K_d values, first the average chemical shift perturbations ($\Delta\delta_{avg}$) at each titration point for 7-10 non-overlapped peaks were calculated:

$$\Delta\delta_{avg} = \sqrt{\left(\frac{\Delta\delta_N}{10}\right)^2 + (\Delta\delta_H)^2}$$

where:

$\Delta\delta_N$ chemical shift change in ^{15}N dimension

$\Delta\delta_H$ chemical shift change in ^{1}H dimension

Then, $\Delta\delta_{avg}$ were plotted as a function of titrant : analyte (depending on the set-up either ligand : protein or protein : peptide) ratio and a two-parameter nonlinear least-squares fit of the data was performed for each peak in the program Origin (OriginLab), using a one-site binding model which corrects for dilution effect (Kannt et al., 1996):

$$\Delta\delta_{avg} = \frac{1}{2}\Delta\delta_{max}\left(A - \sqrt{A^2 - 4R}\right)$$

$$A = 1 + R + \frac{PR + C}{PCK_a}$$

where:

$\Delta\delta_{avg}$ average chemical shift perturbation of a given resonance at a given titration point

$\Delta\delta_{max}$ chemical shift perturbation of a given resonance for $R \rightarrow \propto$

R titrant : analyte ratio

P analyte concentration at the start of the titration

C stock concentration of the titrant

K_a association constant

K_d values are reported as an average ± 2 standard deviations.

It is import to notice that the appearance of NMR spectra of the complexes is affected by the rate of exchange process between the free and bound states of the interacting molecules. Three different exchange regimes can be distinguished:

Slow exchange $2\pi|\nu_F - \nu_B| \gg \left(\frac{1}{\tau_F} + \frac{1}{\tau_B}\right)$

Intermediate exchange $2\pi|\nu_F - \nu_B| \sim \left(\frac{1}{\tau_F} + \frac{1}{\tau_B}\right)$

Fast exchange $2\pi|\nu_F - \nu_B| \ll \left(\frac{1}{\tau_F} + \frac{1}{\tau_B}\right)$

where:

ν_F resonance frequency of the nucleus in the free state

ν_B resonance frequency of the nucleus in the bound state

τ_F lifetime in the free state

τ_B lifetime in the bound state

In case of slow exchange two separate resonances are observed for a given nucleus at frequencies of free and bound species. In case of fast exchange a single averaged resonance is observed at frequency corresponding to the weighted average of frequencies of free and bound species. Thus, in case of fast exchange, the position of the resonance reports on the fraction of the bound molecules and this relationship is used to derive K_d values of the complex using the methodology described above. In case of intermediate exchange, the lineshape of the resonances is not longer Lorentzian and despite a single maximum can still be observed it does not longer report accurately on the fraction of bound molecules. Thus, inclusion of resonances in the intermediate exchange into analysis used to derive K_d values can introduce significant systematic errors (Feeney et al., 1979). In the case of titrations of FIR RRM1-RRM2 with FBP and FBP2 Nboxes the resonances with the highest $\Delta\delta_{avg}$ were in the intermediate exchange as revealed by resonance broadening for partially (~20-80%) saturated protein. One solution to this situation is to exclude such the resonances from the analysis and to perform the analysis for resonances with small difference in the resonance frequencies between the free and bound states. However, in this case analysis relies on measuring small shifts in the peak positions that can be compromised by the resolution of the NMR spectra. Indeed in case of FIR RRM1-RRM2 titrations with FBP/FBP2 Nboxes the values of chemical shift difference were too small to allow a reliable measurement of the dissociation constants. Therefore, we simulated peak shifts and lineshape changes using a macro written by Dr Thomas A. Frenkiel (National Institute for Medical Research, London, UK) based on numerical solution of the Bloch-McConnell equation for the propagation of errors (Figure 9). The simulation was performed using a 15 Hz linewidth for the protein, a K_d of 10 μM, a k_{on} of 1 x 10^8 $M^{-1}s^{-1}$ and a constant protein concentration of 50 μM on three resonances that shift respectively by 20 Hz, 50 Hz and 200 Hz. These conditions recapitulate the ones of our titrations and place the resonance with the larger shift in a fast-to-intermediate regime of exchange. The differences between the K_ds determined (Figure 9B) for the three resonances provide a realistic error for our experimental conditions and indicate that the error we use is larger (see

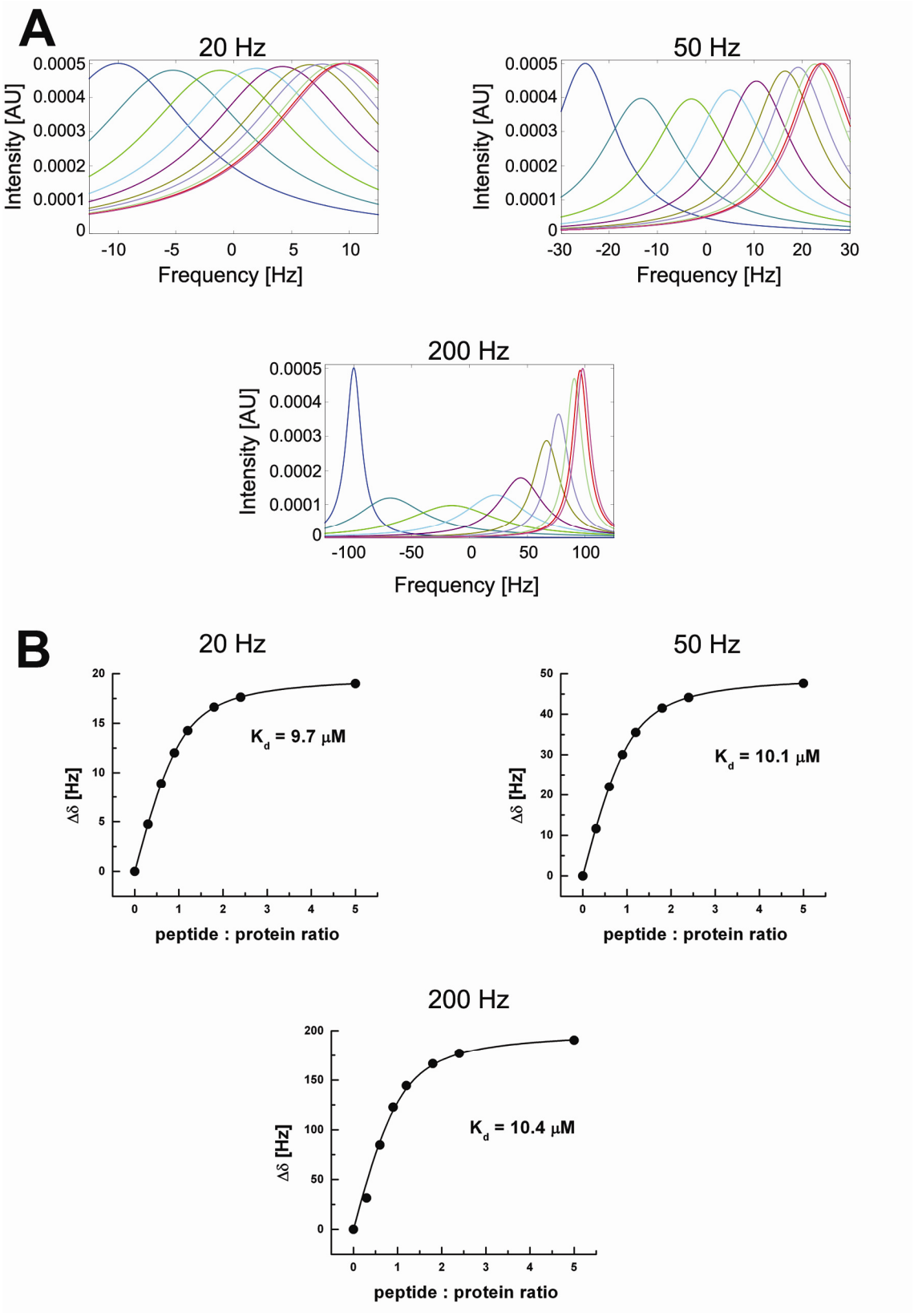

Figure 9 Simulated lineshape changes (A) and peak shifts (B) of three resonances that shift respectively by 20 Hz, 50 Hz and 200 Hz during a titration. The simulations were performed for protein linewidth of 15 Hz, protein concentration of 50 μM, a K_d of 10 μM and a k_{on} of 1 x 10^8 $M^{-1}s^{-1}$.

Figure 33) than the theoretical error due to inclusion of resonances with the highest shifts into the analysis.

2.4.8.2 Scaffold-Independent Analysis

Scaffold-Independent Analysis (SIA) was performed by Dr. Irene Diaz-Moreno (National Institute for Medical Research, London, UK) as described in Beuth *et al.* (Beuth et al., 2007) with 16 pools of 5mer DNA oligonucleotides (nANNN, nGNNN, nCNNN etc., where N is any nucleotide). In brief, solutions of 25 μM ^{15}N-labelled FIR RRM1-RRM2 in 10 mM Tris-HCl pH 7.4, 50 mM NaCl, 2 mM TCEP were titrated with individual DNA pools (ratios 1 : 0, 1 : 1 and 1 : 4). ^{15}N sofast HMQC spectra (Schanda et al., 2005) were recorded at each titration point at 37 °C on a Varian Inova spectrometer equipped with cryoprobe and operating at 800 MHz ^{1}H frequency. Average chemical shift perturbations $\Delta\delta_{avg}$ (calculated as above) of 15 peaks were analysed to obtain SIA scores (Beuth et al., 2007). First, for each peak, the values of $\Delta\delta_{avg}$ at the final titration point with individual DNA pools were calculated. Then, the values of $\Delta\delta_{avg}$ obtained from titrations with four DNA pools with a fixed nucleotide in one position (e.g. for nucleotide fixed in position 1 DNA pools used were nANNN, nGNNN, nCNNN and nTNNN) were normalised to the highest of the four. Finally, for each DNA pool, the normalised shift values were averaged over all 15 peaks to yield a final SIA score for a given nucleotide at a given position. For an example of the oligonucleotide composition of DNA pool and calculation of SIA scores for FIR RRM1-RRM2 protein see Appendix IV. The final results were visualized with WebLogo (http://weblogo.berkeley.edu/logo.cgi) and validated by performing the titration experiments with four ad hoc 5mer ssDNA, which were chosen based on the SIA scores.

2.4.8.3 Bio-Layer Interferometry

All BLI experiments were performed in 10 mM Tris-HCl pH 7.4, 150 mM NaCl, 2 mM TCEP, 0.5 mg/ml BSA on an Octet Red instrument (ForteBio, Inc., Menlo Park, CA) operating at 25 °C. Streptavidin-coated biosensors with immobilised biotinylated ssFUSE DNA (Sigma) were exposed to different concentrations of FIR RRM1-RRM2, FBP Nbox-KH1-KH4 and FBP KH1-KH4 or combinations of them as described in the results section.

The simplest model for the interaction of immobilized DNA (D) with a mobile protein (P) involves the reversible formation of a 1 : 1 complex (DP), with association rate constant k_{on} (units: $M^{-1}s^{-1}$) and dissociation rate constant k_{off} (units: s^{-1}). The equilibrium dissociation constant for the interaction is given by $K_d = k_{off}/k_{on}$ (units: M). Under pseudo first-order conditions ($[P_{Tot}] >> [D_{Tot}]$) the differential rate equation describing the change in concentration of DP with time is:

$$\frac{d[DP]}{dt} = k_{on}[P_{Tot}]([D_{Tot}] - [DP]) - k_{off}[DP]$$

Because the instrument's response, R, is directly proportional to the amount of complex formed this equation can be rearranged to give:

$$\frac{dR}{dt} = k_{on}R_{max}[P_{Tot}] - R(k_{on}[P_{Tot}] + k_{off})$$

where:

R_{max} the response observed at complete saturation of the immobilised DNA

The time course for the association phase is then described by the exponential function:

$$R(t) = R_{eq}\left(1 - e^{-k_{obs}t}\right)$$

where:

R_{eq} the plateau response observed at a particular value of $[P_{Tot}]$

k_{obs} observed rate constant

The observed rate constant k_{obs} (units: s^{-1}) can be determined using standard non-linear least-squares methods and depends on protein concentration according to:

$$k_{obs} = k_{on}[P_{Tot}] + k_{off}$$

The association and dissociation rate constants can be determined by plotting k_{obs} against $[P_{Tot}]$. The ratio of the kinetic constants (k_{off}/k_{on}) is equal to the equilibrium dissociation constant (K_d).

When the sensors are moved to wells without protein (at time t_0) the complex dissociates exponentially with time according to the following equation:

$$R(t) = R(t_0)e^{-k_{off}(t-t_0)}$$

Analysis of the dissociation phase therefore provides an independent measure of k_{off}.

An independent value for the equilibrium dissociation constant can also be determined from the values of the plateau signal (R_{eq}) observed at different protein concentrations using the following equation:

$$R_{eq} = \frac{[P_{Tot}]R_{max}}{K_d + [P_{Tot}]}$$

where:

R_{max} the response observed at complete saturation of the immobilised DNA

This approach can also be applied for reactions that are too fast for a kinetic analysis.

It is important to notice that the values of k_{on}, k_{off} and K_d reported are derived from single experiments because the multiple repetitions were not feasible due to the lack of the material. Therefore the errors reported are the fitting errors and they do not account on the possible variations in the values obtained from independently repeated experiments.

The BLI analysis must also consider that, in principle, the DNA oligonucleotides used are long enough to accommodate an additional protein molecule(s) bound on non optimal site. Indeed at high protein concentration it was necessary to use a double exponential fits to account for a minor component probably associated with some residual non-specific binding to the DNA. Furthermore, in case of FIR protein at least

two optimal DNA sequences are present in the ssFUSE40 DNA used. Due to the lack of additional information, the analysis was performed with the assumption that the multiple binding sites for FIR are equivalent and non-interacting. This is based on the similar affinity that RRM1-RRM2 shows for the two high affinity T and TG-rich sites present on ssFUSE40. The consistency of the results obtained from the analysis of association phase, dissociation phase and the amplitude of BLI signal suggests that the non-specific binding does not significantly affect the results.

Chapter 3

Results – Structure of the first two RNA recognition motifs (RRM) of FIR (FIR RRM1-RRM2)

3.1 Expression and purification of the FIR RRM1-RRM2 protein

The FUSE-mediated regulation of *c-myc* proto-oncogene relies on the interplay between ssFUSE, FBP and FIR. The first two RRMs of FIR (FIR RRM1-RRM2) are necessary and sufficient to interact with ssFUSE and FBP and FBP2, but not FBP3, proteins (Chung et al., 2006). Thus, I undertook further studies to structurally and biophysically characterise this part of FIR protein.

The first challenge towards *in vitro* studies of isolated domains is obtaining stable and soluble protein. Nowadays, many computational methods are available to predict the boundaries of the domains within protein (Card and Gardner, 2005). However, small variations in the positions of N- and C-termini of the expressed constructs may greatly affect the behaviour of the domains in the solution (Card and Gardner, 2005). Therefore, parallel cloning and expression of slightly different constructs are frequently performed to screen for the solubility and stability of the domains (Card and Gardner, 2005). Two forward and two reverse primers were designed upstream and downstream of FIR cDNA sequences coding for RRM1 and RRM2 domains (Table 1, Figure 10). Combinations of these primers were used to clone four variants of each of FIR RRM1, FIR RRM2 and FIR RRM1-RRM2 (Figure 10). The soluble proteins were obtained for FIR RRM1-RRM2 and FIR RRM2 constructs, but not for FIR RRM1 (data not shown).

The ^{15}N-labelled samples of the longest variants of the FIR RRM2 (residues 189-297) and FIR RRM1-RRM2 (residues 103-297) were prepared for further characterisation. The proteins were purified from the soluble fractions of the bacterial cells lysates on IMAC columns. SDS-PAGE analysis revealed bands with the masses corresponding to the HisTag-GST fusion proteins of FIR RRM1-RRM2 (~ 50 kDa) and FIR RRM2 (~ 40 kDa) (Figure 11, Elution 100 lanes). The second intense band (~ 30 kDa) in the FIR RRM2 sample likely represents the fusion itself suggesting that the construct is not well tolerated by *E. coli* cells. The fusions were removed from the recombinant proteins by TEV protease digestion and second IMAC step, yielding the bands with expected molecular masses (~ 22 kDa for FIR RRM1-RRM2 and ~ 10 kDa for FIR RRM2) (Figure 11, Flow-through 1 and 2 lanes). Finally, these samples were cleaned using SEC and stored at -80 °C until use.

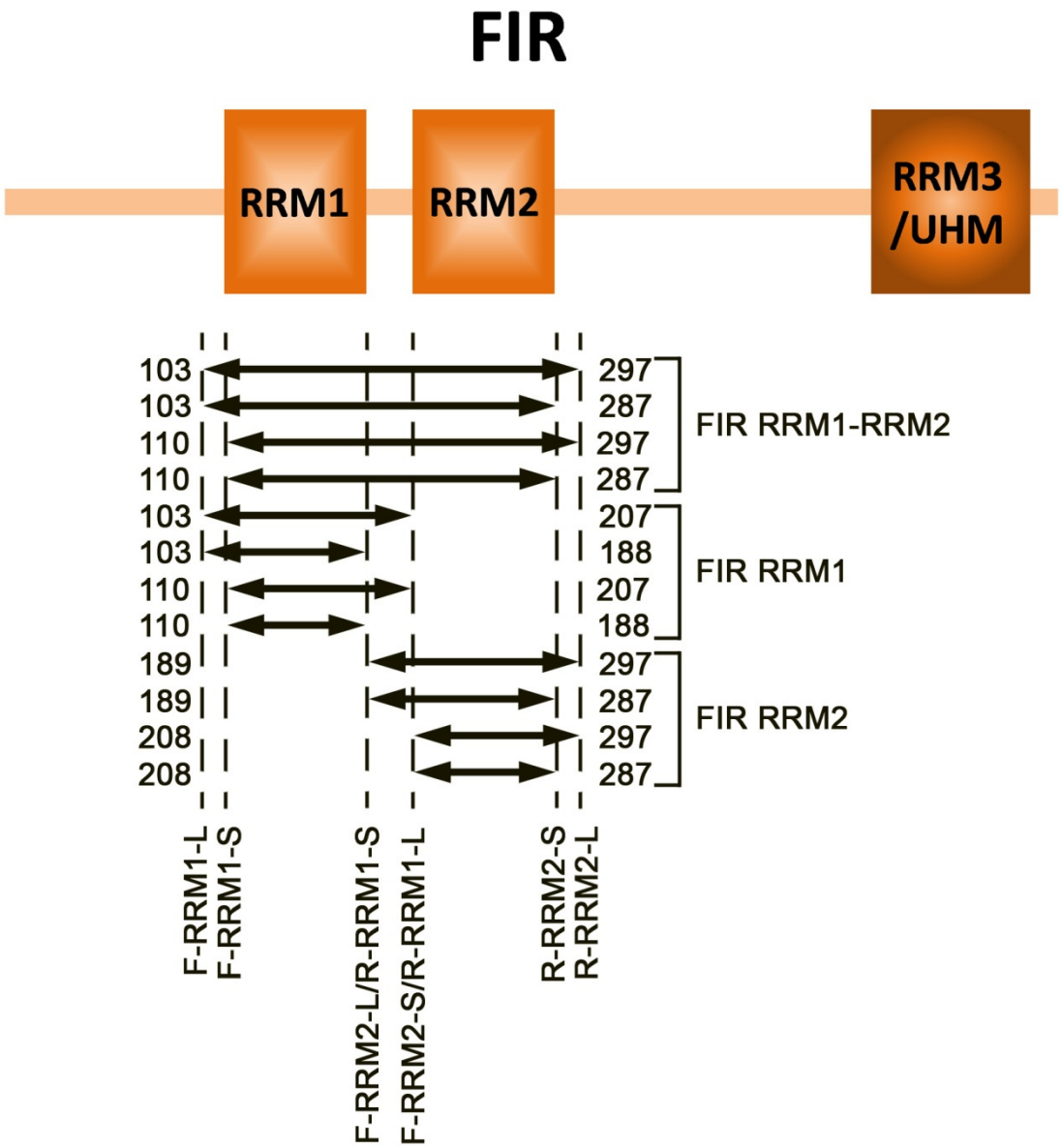

Figure 10 Cloning of the first two RRM domains of FIR. Eight primers (bottom) were used to clone four variants of each of FIR RRM1-RRM2, FIR RRM1 and FIR RRM2 (arrows).

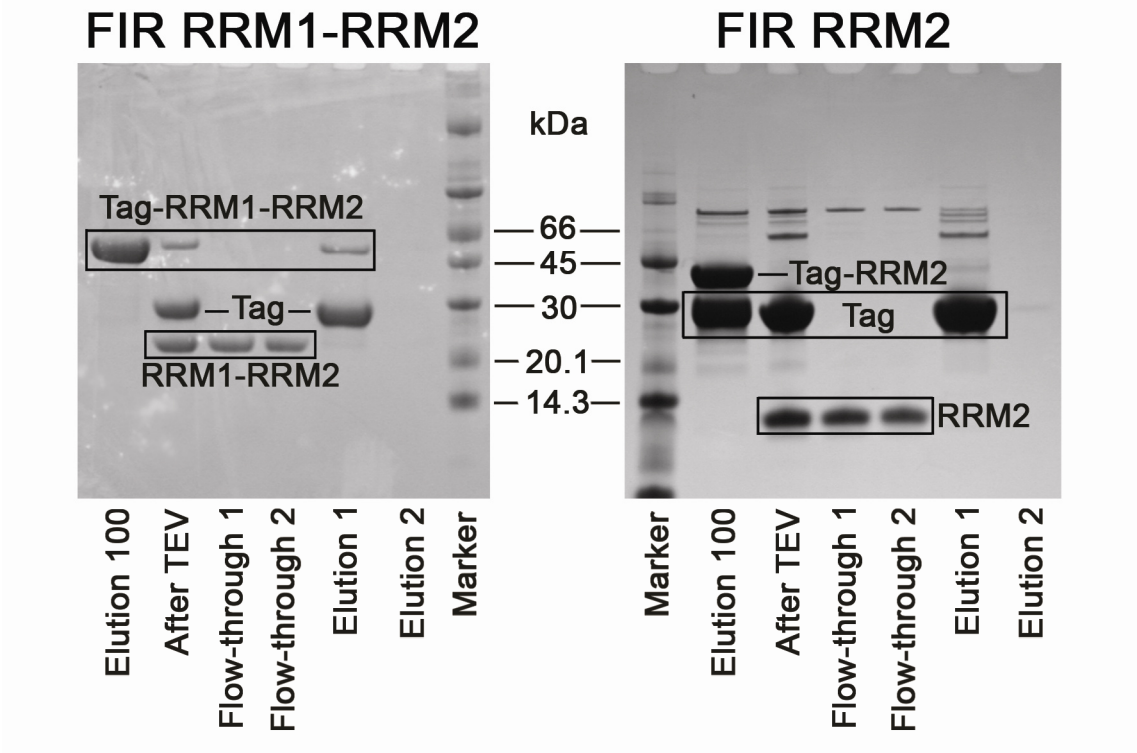

Figure 11 SDS-PAGE analysis of the recombinant FIR RRM1-RRM2 and FIR RRM2 proteins. The recombinant fusion proteins were purified on the IMAC columns from the soluble fractions of the bacterial cell lysates (Elution 100). The eluted proteins were treated overnight with TEV protease (After TEV) and the fusions (HisTag-GST) were removed by a two-step purification on the IMAC columns. The recombinant FIR RRM1-RRM2 and FIR RRM2 were found in the flow-through fractions (1 and 2), while the tags were retained on the columns (Elution 1 and 2).

3.2 Initial characterisation of the FIR RRM2 and FIR RRM1-RRM2 proteins

3.2.1 NMR spectroscopy

The fold and stability of the recombinant FIR RRM2 and FIR RRM1-RRM2 proteins were assessed using ^{1}H-^{15}N correlation NMR experiments, where each peak reports on a pair of covalently bonded ^{1}H and ^{15}N nuclei. The resonance frequencies of the observed nuclei and therefore the position of the corresponding peaks in the spectrum are dependent on the local chemical environments. These resonance frequencies are also dependent on the magnetic field strength at which the NMR measurements are performed and to remove this variable and facilitate the comparison of different experiments at various field strengths the frequencies are reported on a chemical shift parts per million (ppm) scale according to the equation:

$$\delta[ppm] = 10^{6} \cdot \frac{\nu - \nu_{ref}}{\nu_{ref}}$$

where:

δ chemical shift in ppm

ν Larmor frequency of a given nucleus

ν_{ref} Larmor frequency of the line from the agreed reference compound

The influence of chemical environment on the resonance frequencies allows one to evaluate the fold of a protein from the dispersion of the signals in the NMR spectra. In an ^{1}H-^{15}N correlation spectrum, the chemical shifts of the backbone amide groups of proteins in a random-coil conformation are clustered around 125 ppm (except glycines – 110 ppm) in the ^{15}N dimension and 8.3 ppm in the ^{1}H dimension, whereas structured proteins resonances normally show larger signal dispersion (Rehm et al., 2002). 2D sofast ^{15}N HMQC spectra of FIR RRM2 and FIR RRM1-RRM2 recorded at 27 °C show well dispersed signals (~ 106 ppm to ~ 132 ppm in a ^{15}N dimension and ~ 6.5 ppm to ~ 11 ppm in a ^{1}H dimension) (Figure 12) and confirm the folded state of the recombinant proteins. Interestingly, the peaks of FIR RRM2 do not overlap with FIR RRM1-RRM2 signals, showing that amide groups of the same residues in the two

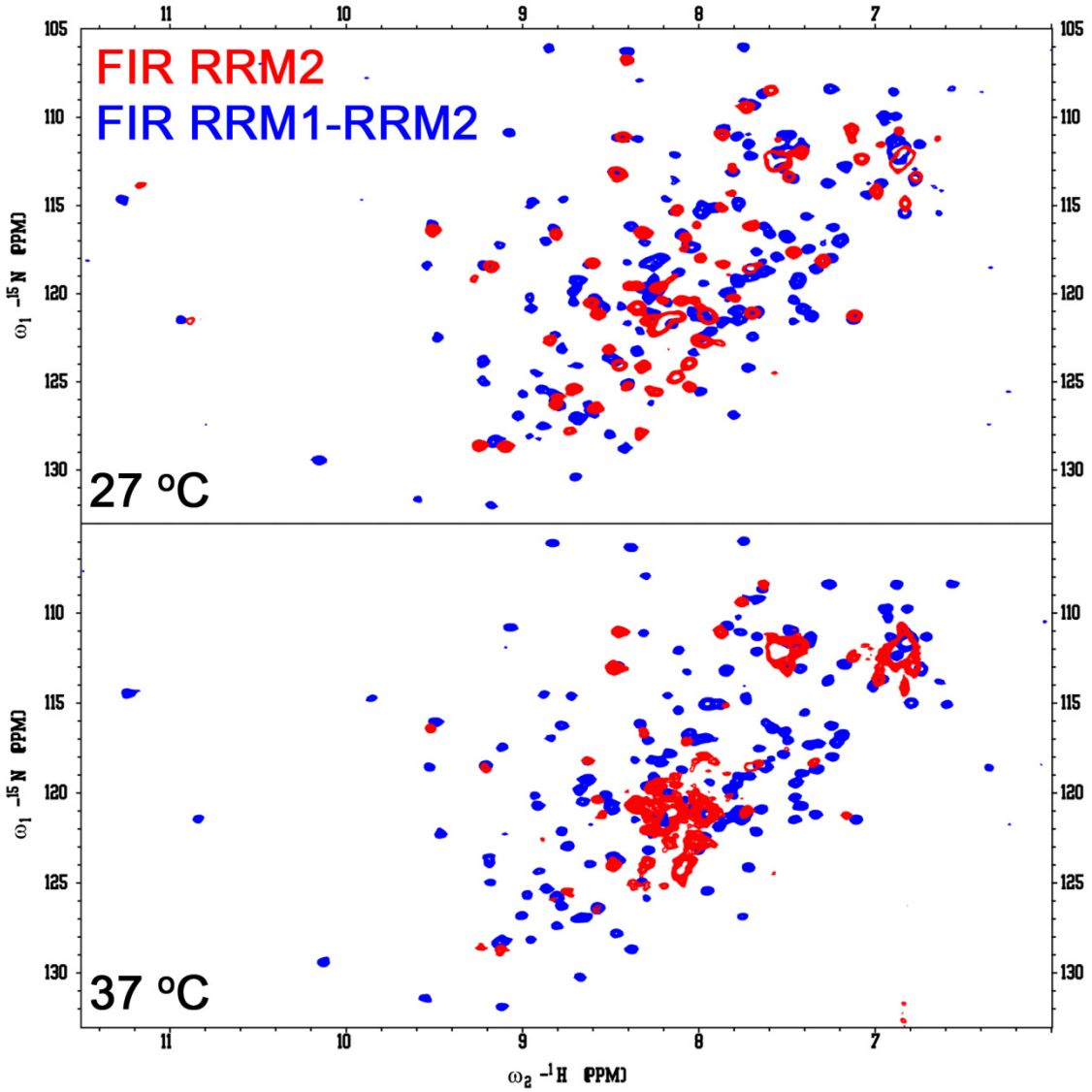

Figure 12 NMR studies of FIR RRM2 (red) and FIR RRM1-RRM2 (blue) fold and stability. 2D sofast ^{15}N HMQC were recorded on 0.1 mM samples in 10 mM Tris-HCl pH 7.4, 50 mM NaCl, 2 mM TCEP at 27 ºC (top) and 37 ºC (bottom). FIR RRM2 interacts with RRM1 and is not stable at higher temperature.

constructs are not in equivalent chemical environments and confirming that FIR RRM1 and RRM2 interact with each other. Furthermore, at 37 °C FIR RRM1-RRM2 is still properly folded, while FIR RRM2 equilibrium is shifted towards unfolded species, demonstrating that its stability depends on the interaction with RRM1 (Figure 12).

3.2.2 Circular dichroism

The far-UV CD spectrum of a protein reflects the secondary structure content (Figure 13A, inset) and the spectra of proteins that contain different secondary elements are assumed to be a linear combination of the spectra of the individual elements (Martin and Schilstra, 2008). Importantly, the contributions are not equivalent because of differences in the CD extinction coefficients of chromophores within different secondary structure elements, for example in case of αβ proteins (e.g. RRM domains) the spectrum is dominated by the α-helical component (Martin & Schilstra, 2008). The far-UV spectrum of the FIR RRM1-RRM2 protein shows the presence of α-helices (negative bands at ~ 208 nm and ~ 222 nm) and confirms the folded state of the protein (Figure 13A).

The CD signal changes in response to conformational rearrangements and therefore it can be used to study the thermal stability of the protein. A sample of FIR RRM1-RRM2 was gradually heated and CD signal was monitored at 220 nm. The data showed that thermal unfolding of RRM1-RRM2 protein is irreversible and therefore a full thermodynamic analysis is not possible. However, the transition midpoint of the protein, defined as the temperature at which a CD signal is halfway between the values for the two states (that I assume to be folded and unfolded protein), is 57.2 °C (Figure 13B). Interestingly, a single-step transition is observed, indicating that RRM1 and RRM2 unfold at similar or the same temperature, which is in agreement with the inter-domain interaction.

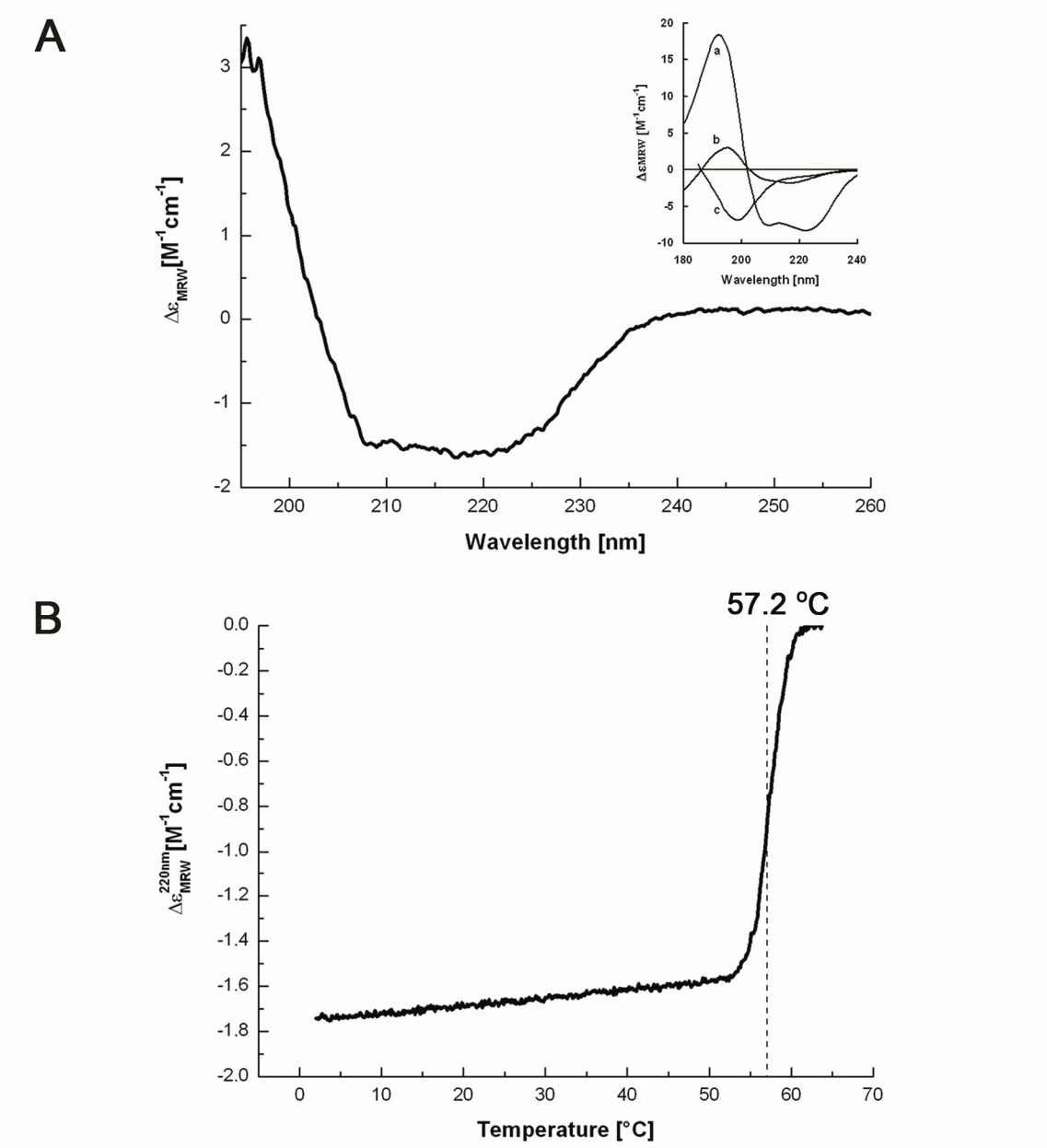

Figure 13 CD studies of the FIR RRM1-RRM2 protein fold and stability. A) Far-UV spectrum of the RRM1-RRM2 recorded at 25 °C and pH 7.4. The minima at 208 nm and 222 nm indicate a significant α-helical content. Inset – the far-UV spectra of: a) – an all-α-helical protein (myoglobin, 4mbn); b) – an all-β-strand protein (prealbumin, 2pab) and c) – an unordered protein (acid denaturated staphylococcal nuclease). All spectra are available on the internet at: http://lamar.colostate.edu/~sreeram/CDPro/. B) CD-monitored thermal unfolding curve of the FIR RRM1-RRM2 protein. The dashed line indicates the midpoint of the transition.

3.3 Optimisation of the experimental parameters for NMR studies

The S/N ratio of NMR signal is dependent on the concentration of the protein and its molecular weight/aggregation state that affect the resonance linewidth. The resonance linewidth, defined as the width at half-height of the peak (Figure 14), is proportional to the transverse relaxation rate constant R_2. The values of the R_2 rate constant, in turn, are proportional to the overall rotational correlation time (τ_c) that reports on the molecular weight and shape of the studied species. For a globular non-aggregated protein, a theoretical values of τ_c can be obtained from Stoke's equation (Cavanagh, 2007):

$$\tau_c = \frac{4\pi\eta_w r_H^3}{3k_B T}$$

where:

η_w viscosity of the solvent

k_B Boltzmann constant

T temperature

r_H effective hydrodynamic radius of the protein, that can be estimated from the molecular weight of the protein from the equation:

$$r_H = \sqrt[3]{\frac{3\bar{V}M_r}{4\pi N_A}} + r_w$$

where:

$\bar{V}$ specific volume of the protein

M_r molecular weight of the protein

N_A Avogadro's number

r_w thickness of the hydration layer

In general, for the bigger proteins (larger M_r), the R_2 values are higher and consequently, the resonance linewidths are broader. In case of aggregation, the M_r of the species in the sample increases, causing broadening of the resonances and decreasing S/N (Figure 14).

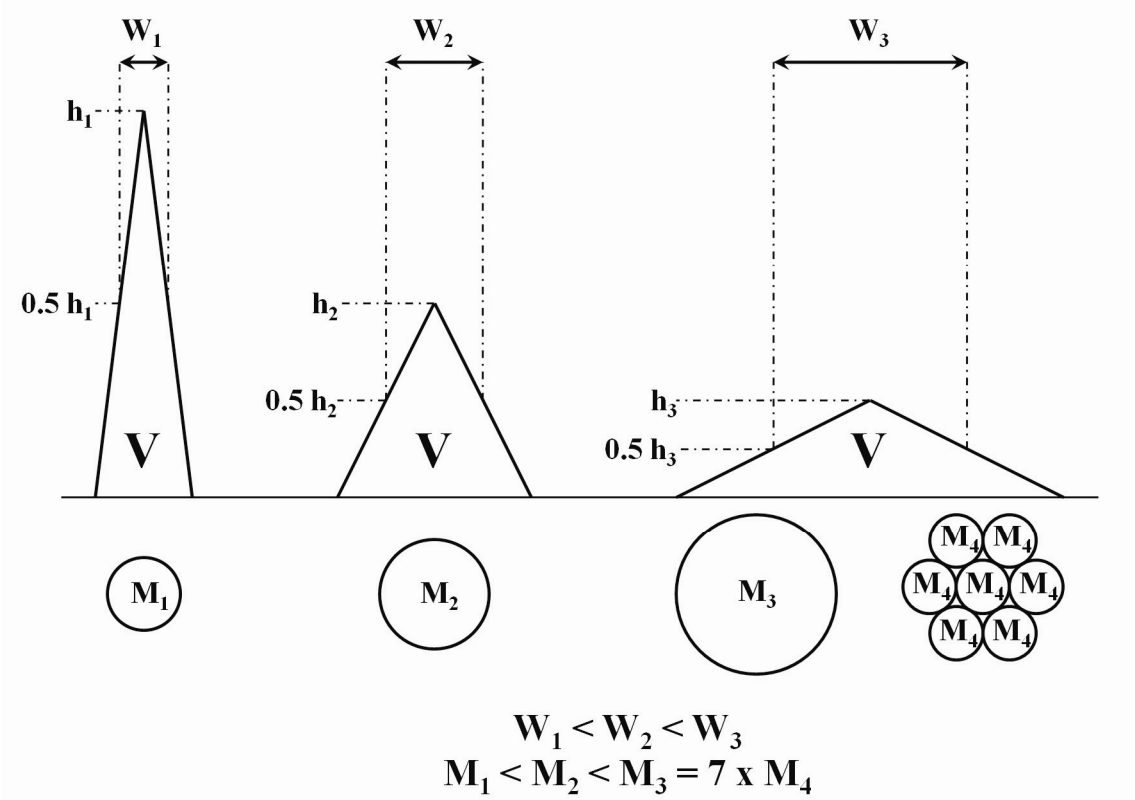

Figure 14 Schematic representation of the relationship between resonance linewidth (W) and molecular weight (M). Resonance linewidth (W) is defined as a full-width at half-height (0.5 h) of the resonance lineshape. Increase in the molecular weight (M) of the species, also due to aggregation (7 x M_4), results in the broadening of resonance linewidth (W). Because the integral of the resonance (V) is constant for the same amount of atoms, the S/N decreases for heavier species.

Initially, it was noticed that the purity of recombinant FIR RRM1-RRM2 is satisfactory prior to SEC (see Figure 11). Therefore, this purification step was omitted to shorten the preparation procedure and to increase the yield of the protein. However, a comparison of the 1D (Figure 15A) and 2D (data not shown) NMR spectra of FIR RRM1-RRM2 (with and without SEC purification) revealed that incorporation of the SEC step improved S/N in the NMR spectra and resulted in narrower linewidths. In the chromatogram from SEC, FIR RRM1-RRM2 protein is eluted at the expected retention volume of ~ 80 ml (Figure 15B). Faster-migrating peaks (at ~ 50 ml, ~ 67 ml and ~ 75 ml) with A_{260}/A_{280} ratio ≈ 2 likely represent trace amounts of nucleic acids co-purified with the protein. We concluded that the presence of nucleic acid broadens FIR RRM1-RRM2 resonances.

The interaction between two macromolecules changes the chemical environment of amide groups at the contact interfaces and therefore ^{1}H-^{15}N correlation experiments could be employed to monitor the interactions in the FUSE–FBP–FIR system. As a preliminary step to the binding assays I optimised quality of the 2D ^{15}N HSQC spectra by screening temperature and salt concentration. The best conditions (37 °C and 50 mM NaCl) were used in subsequent binding experiments where ^{15}N-labelled FIR RRM1-RRM2 protein was titrated with either the FBP Nbox peptide or the 29mer ssFUSE (ssFUSE29) (Figure 16). The selective chemical shift perturbations (CSP) indicate that both the peptide and the ssFUSE bind to the FIR RRM1-RRM2 protein. Based on these results further NMR studies were planned to determine the high resolution structure of the FIR RRM1-RRM2 alone and in the complex with FBP Nbox peptide.

3D backbone NMR experiments used for structural studies of proteins are intrinsically insensitive due to the rapid relaxation rates (high R_2) of the nuclei involved in the magnetisation transfer (Gardner and Kay, 1998). The concentration of protein used must therefore be higher than the one used in ^{1}H-^{15}N correlation experiments (Markley et al., 2003). Unfortunately, preliminary NMR experiments performed on 1 mM sample of $^{15}N^{13}C$-labelled FIR RRM1-RRM2 protein showed that the conditions used for the binding studies, where the protein was at a lower (50-150 μM) concentration, are not suitable to obtain good quality data, probably because of transient aggregation. Therefore, a new set of optimisation experiments was carried out testing a range of protein concentrations, buffers, pH, salt concentrations, anti-aggregation agents and reducing agent concentrations. In most cases no significant change was

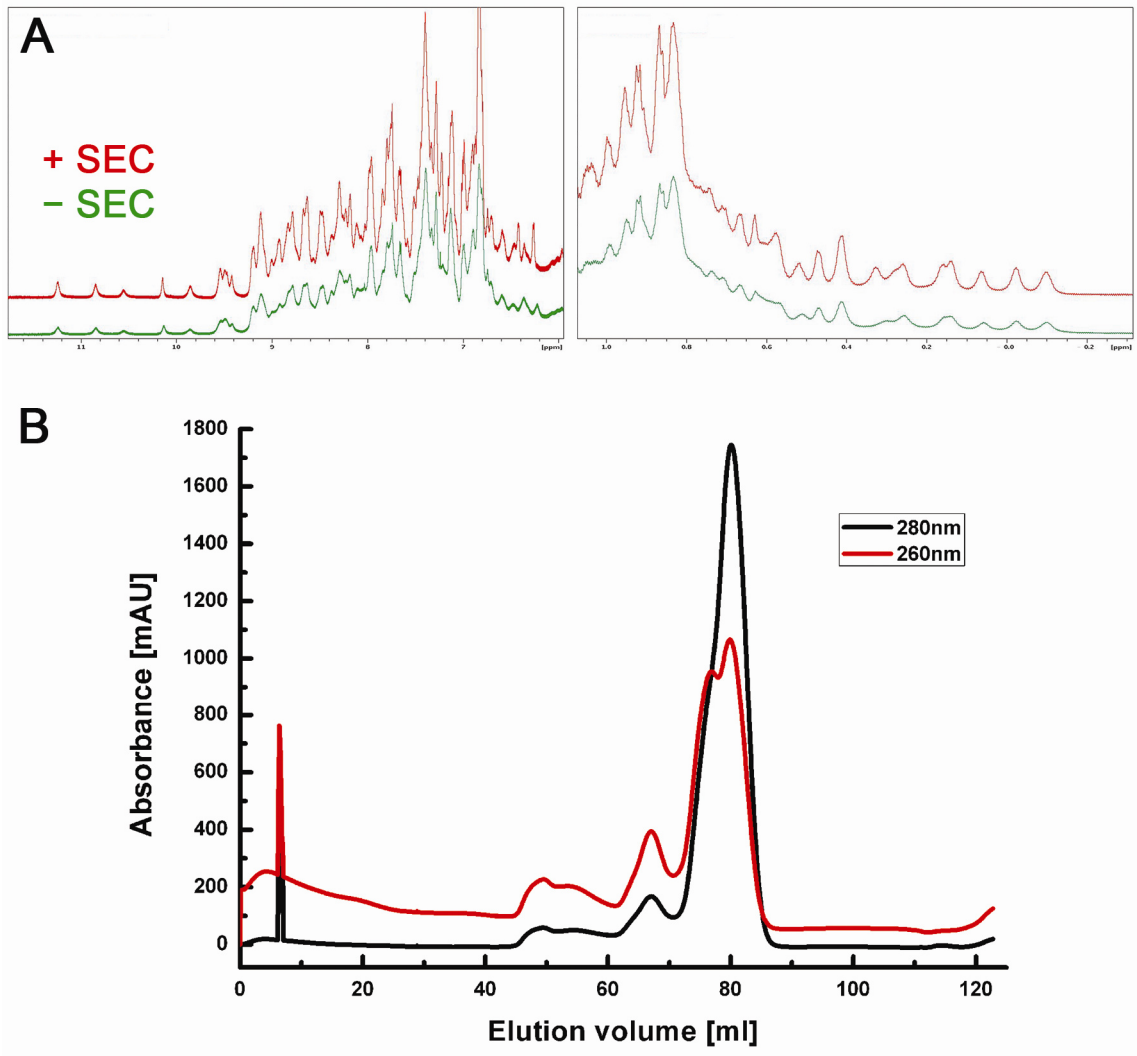

Figure 15 SEC purification step significantly improves S/N. A) Amide proton (left) and methyl (right) regions of 1D ^{1}H NMR spectra recorded on 0.35 mM samples of the FIR RRM1-RRM2 that was not (green) or was (red) subject to SEC. Improved S/N and resolution are obtained after SEC purification as a result of narrower resonance linewidths. B) Chromatogram from SEC of the FIR RRM1-RRM2 sample. Faster-migrating peaks with relatively high absorbance at 260 nm are separated from the main protein peak (~ 80 ml).

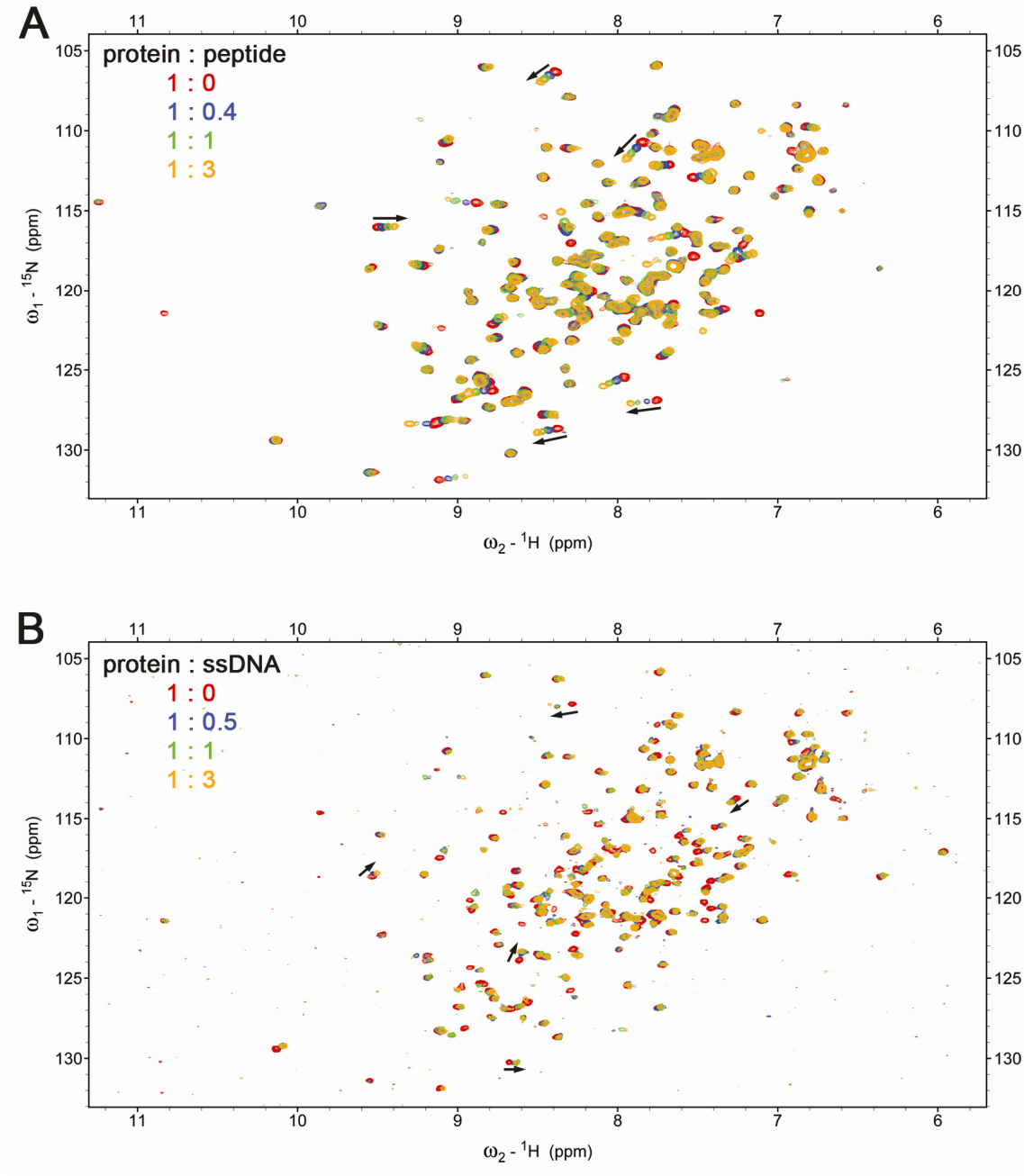

Figure 16 Titrations of the ^{15}N-labelled FIR RRM1-RRM2 with the FBP Nbox peptide and with the ssFUSE29. A) Superimposed 2D ^{15}N HSQC spectra recorded during the titration of the RRM1-RRM2 with the FBP Nbox peptide. The different titration points (protein : peptide ratios) are colour-coded as follow: 1 : 0 – red; 1 : 0.4 – blue; 1 : 1 – green and 1 : 3 – orange. B) Superimposed 2D ^{15}N HSQC spectra recorded during the titration of the RRM1-RRM2 with the ssFUSE. Different titration points (protein : DNA ratios) are colour-coded as follow: 1 : 0 – red; 1 : 0.5 – blue; 1 : 1 green and 1 : 3 orange. In both panels the arrows indicate examples of perturbed cross-peaks (from the free to the bound position).

observed, but ~ 2 fold improvement in S/N was achieved in the 2D (^{1}H-^{13}C) version of the HNCACB spectrum when the protein concentration was reduced from 1 mM to 0.4 mM and the pH was changed from 7.4 to 8.0 (data not shown).

3.4 Backbone and side-chain assignment of the FIR RRM1-RRM2

A ~ 60-70% completeness of backbone assignment was obtained from spectra recorded on ^{15}N^{13}C-labelled FIR RRM1-RRM2 samples, with the missing resonances mainly from residues in RRM1. To obtain a more complete backbone assignment it was necessary to deuterate FIR RRM1-RRM2. Because backbone experiments correlate exchangeable protons with ^{15}N and ^{13}C nuclei, it is possible to substitute the non-exchangeable (^{13}C-attached) protons with deuterons without any loss of information. This substitution removes the dipolar interaction between the ^{13}C nuclei and the attached protons, which represents the largest component of ^{13}C relaxation (Venters et al., 1996). Thus, deuteration leads to the decrease in the rate of transverse relaxation of ^{13}C and ^{1}H^{N} nuclei and to the increase in the sensitivity of multi-dimensional NMR experiments (Gardner and Kay, 1998; Venters et al., 1996). A 60-70% deuteration of the FIR RRM1-RRM2 (as assessed from 1D NMR experiments) provided ~ 2 fold gain in S/N in NMR experiments compared to ^{15}N^{13}C-labelled RRM1-RRM2 at pH 8.0 (data not shown). In summary, the optimisation of the experimental conditions and the deuteration of the sample increased significantly the number of correlations visible in 3D NMR experiments (e.g. HNCACB, Figure 17).

By integrating the information from the different backbone experiments I could assign 160 out of 185 backbone amides of the FIR RRM1-RRM2 (Figure 18, see Appendix V). Interestingly, if the domains are analysed separately the assignment of RRM1 is less complete than the one of RRM2 (59 amino acids out of 72 assigned – 82% completeness for RRM1 *vs*. 72 amino acids out of 76 – 95% completeness for RRM2) (Figure 18). The process of backbone assignment was validated by comparing the secondary structure elements identified based on the chemical shifts of $^{13}C^{\alpha}$ and $^{13}C^{\beta}$

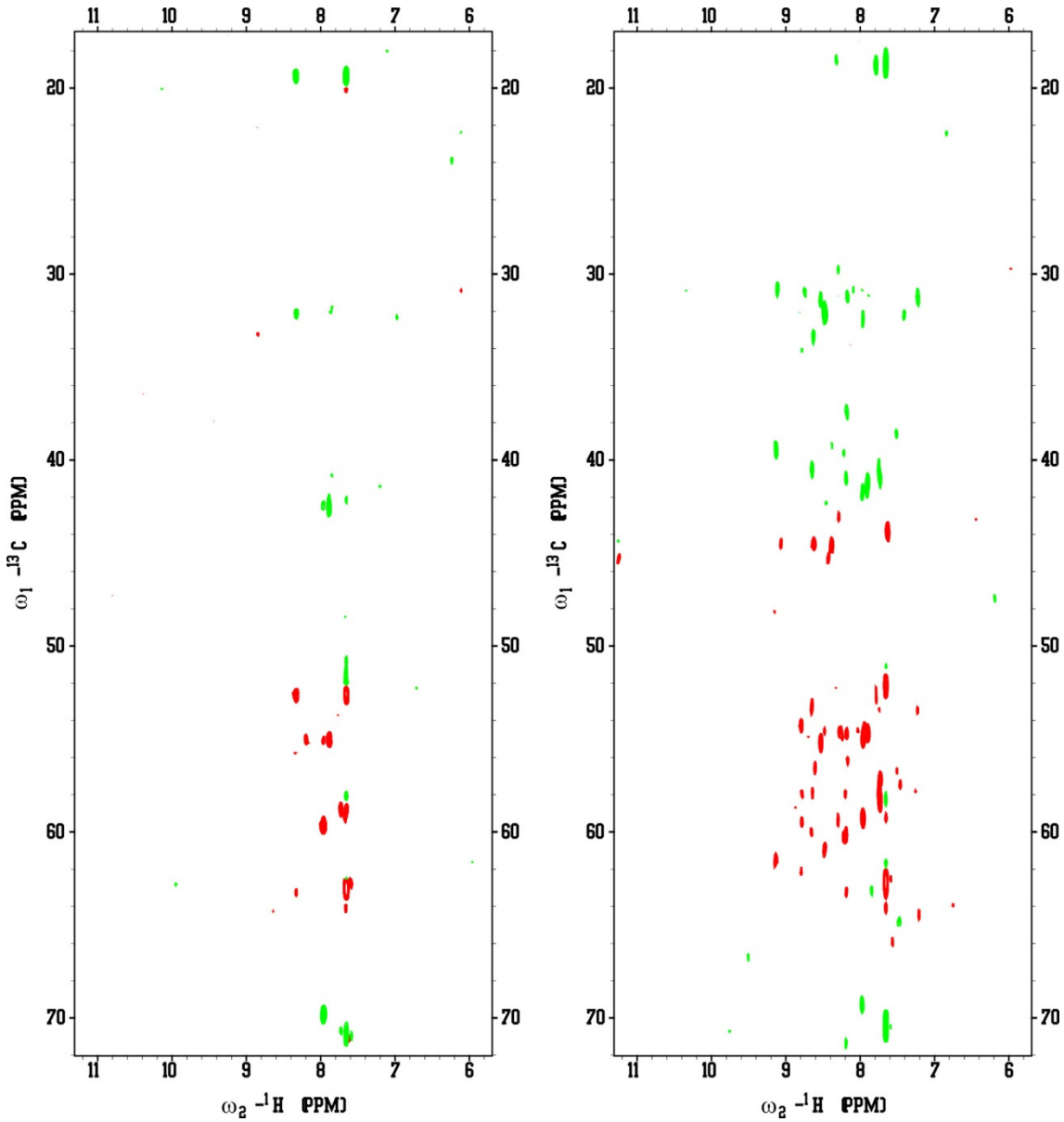

Figure 17 An improvement in sensitivity of through-bond experiments is obtained upon optimisation of the experimental parameters and partial deuteration of the protein. Comparison of the 2D versions of the 3D HNCACB experiment recorded on 1 mM $^{15}N^{13}C$-labelled FIR RRM1-RRM2 at pH 7.4 (left panel) and 0.4 mM $^{15}N^{13}C^{2}H$-labelled FIR RRM1-RRM2 at pH 8.0 (right panel). Only signals of the amino acids in a random-coil region are visible in the 2D HNCACB spectrum recorded on a non-deuterated sample with non-optimised conditions (left panel). Conversely, signals of the amino acids in structured regions, that have significantly higher rotational correlation time, are visible in an equivalent spectrum recorded on the deuterated sample with optimised conditions (right panel).

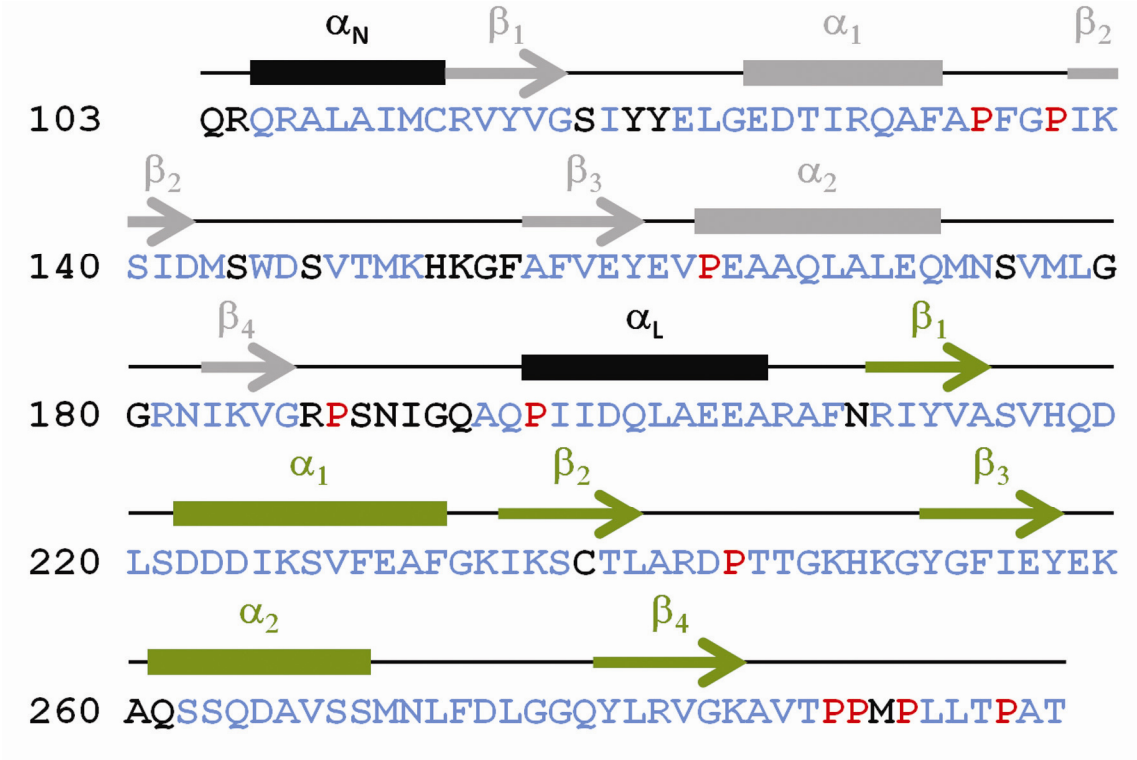

Figure 18 Overview of the backbone assignment of FIR RRM1-RRM2. The residues with assigned and unassigned backbone amide resonances are in blue and black, respectively. Prolines, which are not detectable in ^{1}H-^{15}N correlation spectra, are in red. The N-terminal four residues, that are a by-product of cloning, are not displayed. α-helices (boxes) and β-strands (arrows), identified based on $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts, are indicated above the protein sequence in grey (RRM1), green (RRM2) and black (linkers).

resonances (Metzler et al., 1993) and the characteristic patterns of the NOE cross-peaks (Neuhaus and Williamson, 2000) with the expected arrangement of secondary structure elements within RRM (Maris et al., 2005) (Figure 18). Interestingly, two additional α-helical regions were identified, which are located in the N-terminus of the construct (α_N) and in the linker between the RRMs (α_L).

Side-chain assignment resulted in the identification of most of the ^{1}H, ^{13}C and some ^{15}N resonances (Appendix V). Here, the missing chemical shift assignments are mainly of the residues with long side-chains (e.g. lysines, glutamates, glutamines) located on the protein surface and of the residues in flexible loops that are very overlapped, have often chemical shifts with random-coil values and show broadened methylene resonances most likely due to the high mobility of these regions. A challenging task was also assignment of ^{1}H and ^{13}C in aromatic rings, especially those of phenylalanines (nine in FIR RRM1-RRM2) and tyrosines (eight in FIR RRM1-RRM2). The standard ^{13}C NOESY and HCCH TOCSY experiments used for aliphatic side-chain assignment are not efficient in magnetisation transfer in the aromatic ring. Unfortunately, the large number of aromatic residues made it impossible to assign unambiguously the resonances of ^{1}H from the aromatic ring without further information on the connected ^{13}C resonances. The latter was obtained from a 3D ^{13}C NOESY experiment optimised for aromatic residues recorded on a doubly labelled sample. Furthermore, the increased sensitivity of this experiment for correlations involving ^{1}H and ^{13}C resonances from the aromatic ring was particularly useful in the determination of the structure, because the additional correlations to ring protons fix the position of bulky and rigid aromatic ring in the hydrophobic core of the protein and impose strong restraints on the calculated structures.

3.5 Structure calculations of the FIR RRM1-RRM2

The restraints for structure calculations were derived from 1) NOESY spectra (NOE distance restraints), 2) ^{2}H exchange experiment (H-bond distance restraints) and 3) analysis of chemical shifts with TALOS software (torsion angle restraints). During first run of ARIA most of the intra-residual, some of the sequential and medium-range,

and approximately ~ 300 long-range NOE distance restraints were used as unambiguously assigned, while the rest of NOE cross-peaks were left for automatic assignment by ARIA. These distance restraints were supported with TALOS-derived torsion angle restraints that were available for most of the secondary structure elements. H-bond restraints were not introduced in the first run, because the acceptor group of the H-bond can be identified only indirectly from the examination of preliminary structures.

In the following ARIA runs, I gradually increased the number of restraints and constantly evaluated them to identify and remove/correct any erroneous data. Briefly: 1) I verified the NOE restraints that were unambiguously and automatically assigned by ARIA and, if correct, I used them as manually assigned in the following runs. 2) I inserted an H-bond distance restraint only when an exchange-protected $^{1}H^{N}$ formed an H-bond in at least 50% of the calculated conformers in the preceding ARIA run and this H-bond is consistent with other available restraints. 3) I removed torsion angle restraints from the following runs if the restraint was violated or a large number of NOE distance restraints violations was observed nearby. In the final run of ARIA 4368 NOE distance restraints (2780 inter-residual), 63 H-bond restraints and 236 φ and ψ torsion angle restraints were used (Table 5).

3.6 Solution structure of the FIR RRM1-RRM2

The obtained solution structure of the FIR RRM1-RRM2 is well defined with the root mean square deviation (RMSD) from the mean structure of 0.91 Å for backbone atoms and 1.36 Å for all heavy atoms (Table 5 and Figures 19, 20A and 20B). The exceptions are the loop regions where some flexibility is observed. The domains (RRM1 – residues R113-R187 and RRM2 – residues R210-P288) have the $\beta_1\alpha_1\beta_2\beta_3\alpha_2\beta_4$ topology typical of RRM domains with two α-helices packed against four-strand anti-parallel β-sheet (Maris et al., 2005). The β-sheet is arranged in $\beta_4\beta_1\beta_3\beta_2$ order when looking from the front and with N-terminus of β_1 facing north. Also α_2/β_4 loops of the domains form additional small two-stranded anti-parallel β-sheets as often found in RRMs (Maris et al., 2005).

Table 5 NMR statistics for FIR RRM1-RRM2.

	FIR RRM1–RRM2 (103-297)
NMR distance and dihedral restraints	
Distance restraints	4368
Intra-residue	1588
Inter-residue	2780
Sequential ($\|i - j\| = 1$)	906
Medium-range ($1 < \|i - j\| \leq 4$)	487
Long-range ($\|i - j\| > 4$)	1387
TALOS φ and ψ restraints	236
Hydrogen bond restraints	63
Structure statistics	
Mean total energy (kcal/mol)	-6489 ± 115
Dihedral angles violations > 5º	0
NOE violations > 0.5Å	0
NOE violations > 0.3Å	0.9 ± 0.9
Mean NOE energy (kcal/mol)	62 ± 5
RMSD from idealised covalent geometry	
Bond lengths (Å)	0.0036 ± 0.0001
Bond angles (º)	0.480 ± 0.011
RMSD from the mean structure (Å)	res. 105-288
Backbone atoms	0.91 ± 0.10
Heavy atoms	1.36 ± 0.09
RMSD from the mean structure (Å)	secondary structure elements only*
Backbone atoms	0.59 ± 0.09
Heavy atoms	1.03 ± 0.09
Ramachandran plot analysis	
Most favoured regions (%)	83.9
Additional allowed regions (%)	14.3
Generously allowed regions (%)	0.9
Disallowed regions (%)	0.9

* residues: 105-111, 113-117, 125-132, 138-143, 155-160, 164-172, 184-185, 194-208, 210-214, 222-229, 235-243, 248-257, 260-269 and 281-284

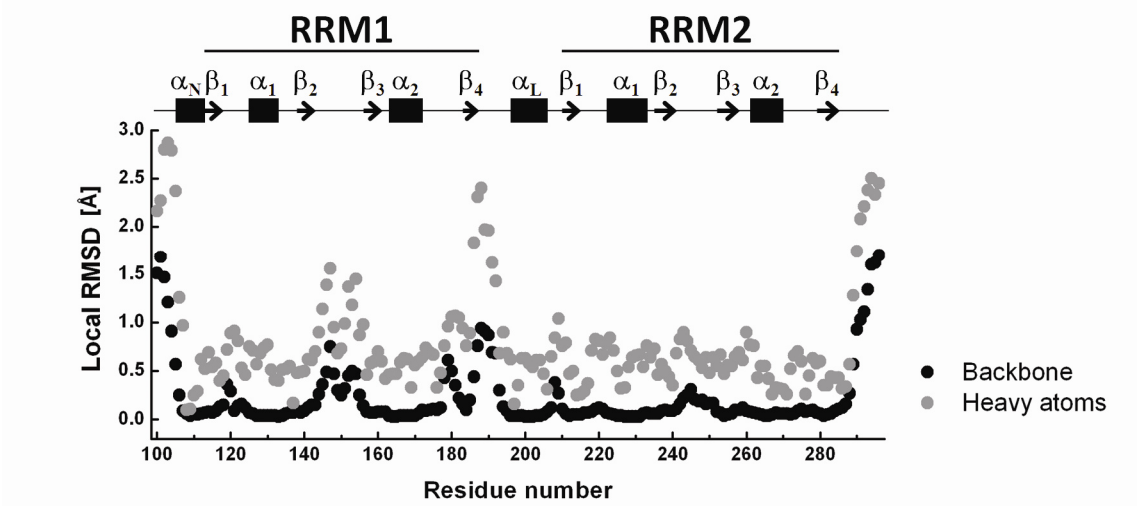

Figure 19 Local RMSD *vs.* FIR RRM1-RRM2 sequence. The 20 calculated conformers were superimposed in Molmol (Koradi et al., 1996) using residues 105-111, 113-117, 125-132, 138-143, 155-160, 164-172, 184-185, 194-208, 210-214, 222-229, 235-243, 248-257, 260-269 and 281-284 and the local RMSD were calculated for backbone and heavy atoms. The highest RMSD are observed at the N- and C-termini and in the loop regions.

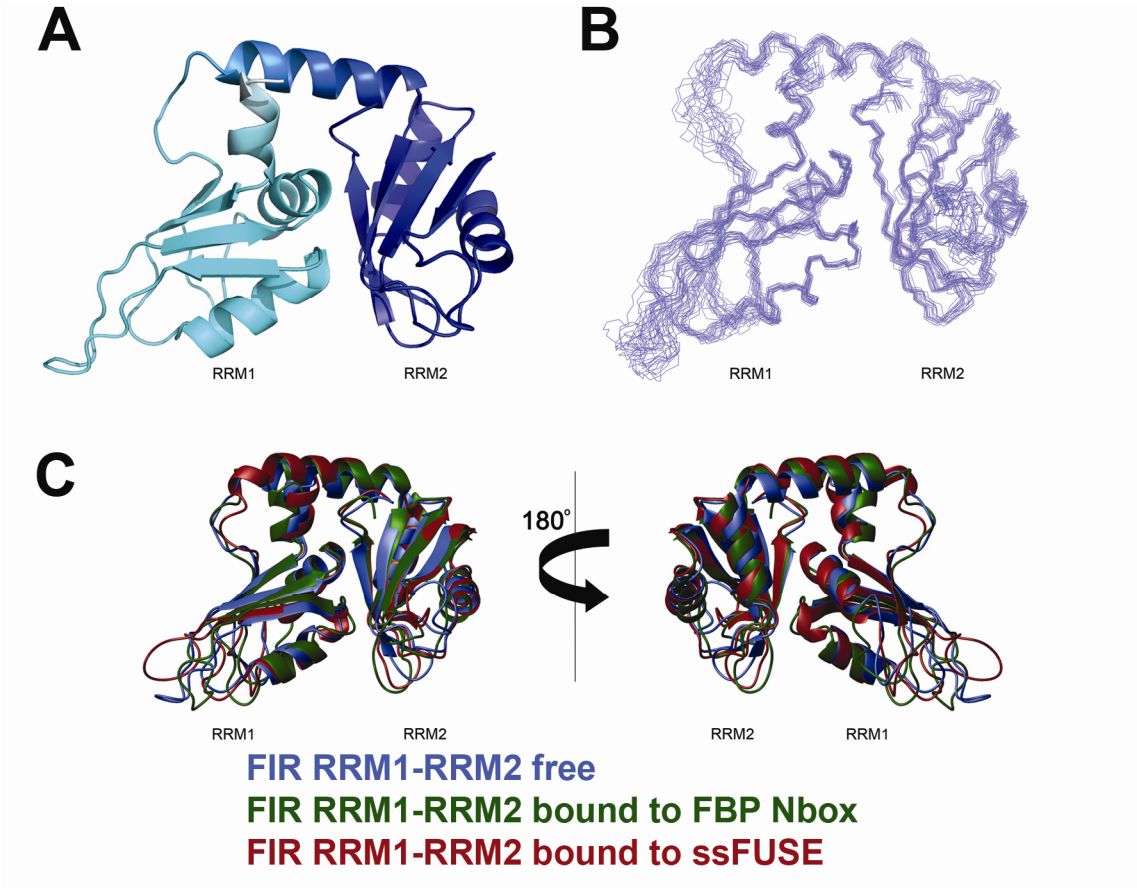

Figure 20 Structure of the FIR RRM1-RRM2. A) Ribbon representation with RRM1 in light blue and RRM2 in dark blue. The amino terminal and linker alpha helices (α_N and α_L) are coloured in white-to-light blue and a light blue-to-dark blue colour gradients, respectively. B) Backbone traces of the 20 lowest-energy FIR RRM1-RRM2 conformers. The structure is well defined, except for a few flexible loops. C) Superimposition of the ribbon representations of the structures of FIR RRM1-RRM2 free (blue, this study), in complex with the FBP Nbox (green, this study) and in complex with ssFUSE (brown, Crichlow et al., 2008). The lowest-energy NMR structures (free and FBP Nbox complex) are shown. There are no major structural changes in the FIR RRM1-RRM2 upon FBP or ssDNA binding.

FIR RRM1 and RRM2 interact with each other through a hydrophobic core formed by α_2 helix and α_1/β_2 loop of RRM1 and the β-sheet of RRM2 (1100 $Å^2$ of buried surface) (Figures 20A and 21). The central part of the interface is occupied by conserved aromatic residues of RRM2: Y212, Y252 and F254 that interact with E164, P137 and V162 in RRM1 (Figure 21A). The inter-domain orientation is fixed by a network of mainly hydrophobic contacts between RRM1, RRM2 and the two additional α-helical elements, e.g. L108 (α_N) – P163 (RRM1), A107 (α_N) – I197 (α_L), L108 (α_N) – T287 (RRM2), P134 (RRM1) – R281 (RRM2) and L201 (α_L) – P288 (RRM2) (Figure 21B).

The recorded relaxation data confirmed that RRM1 and RRM2 do not tumble independently giving a rotational correlation time τ_c of 10.6 ns, as expected for a ~ 22 kDa single-unit protein (Maciejewski et al., 2000). Furthermore, the steady-state heteronuclear NOE (hetNOE) values clearly indicate that flexibility is limited only to the loop regions, particularly β_2/β_3 loops (Figure 22). Overall FIR RRM1-RRM2 structure is very similar in the free form and in the complex with FBP Nbox (discussed later) or 25mer ssFUSE (Crichlow et al., 2008) (Figure 20C). The calculated RMSD between an averaged NMR structure in the free form and the X-ray structure is 1.9 Å for backbone atoms and 2.4 Å for all heavy atoms. Furthermore, the relaxation data show that there is no major variations in the internal dynamics upon binding of FIR RRM1-RRM2 to FBP Nbox or ssDNA, except expected increase in the rotational correlation time and decreased T2 values at the protein – peptide interface that will be discussed later (Figure 22).

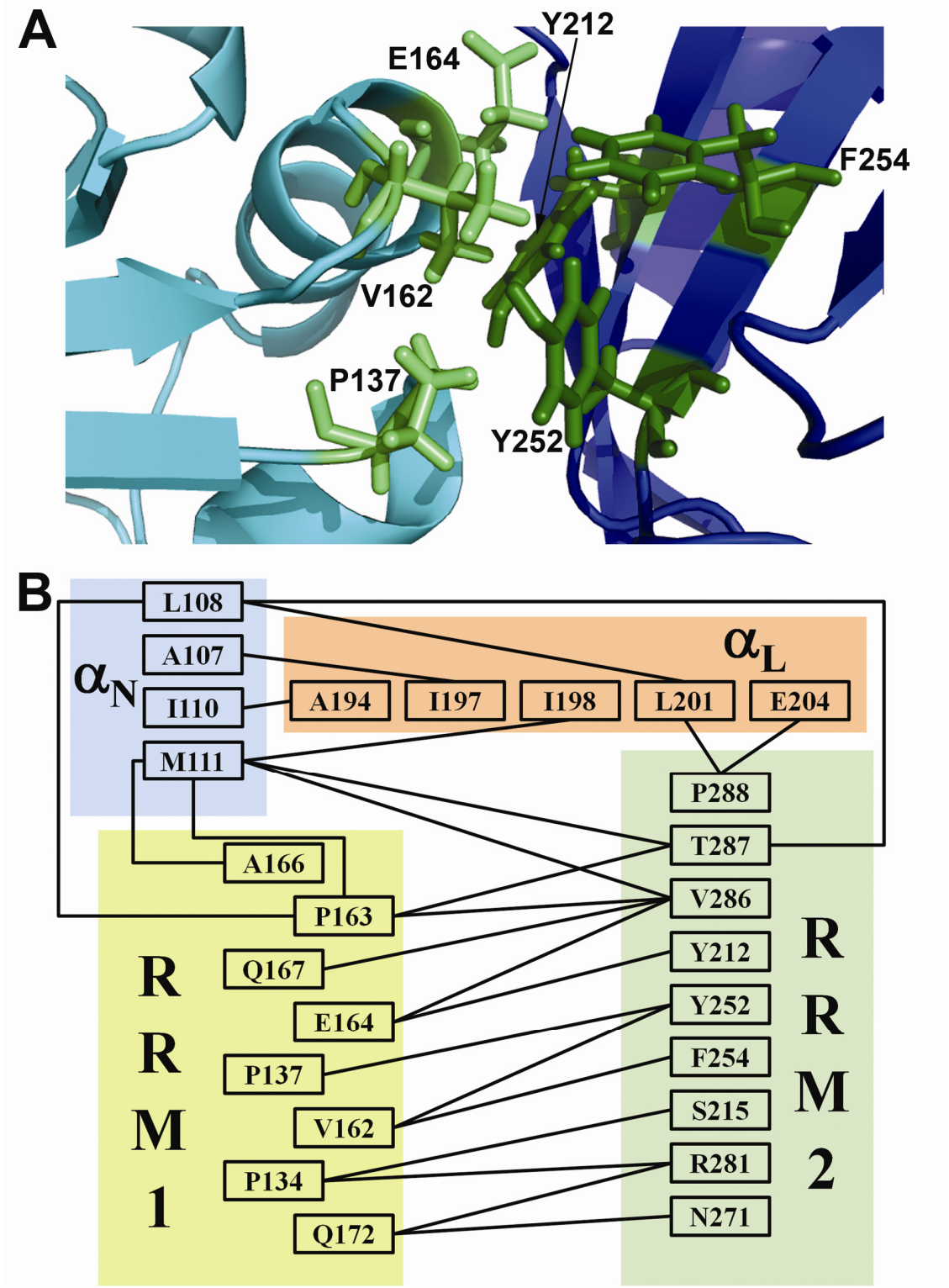

Figure 21 The inter-domain interaction between RRM1 and RRM2 has a hydrophobic character. A) Detailed view of a core hydrophobic inter-domain interaction. RRM1 is in light blue and RRM2 in dark blue. The key residues are displayed in a stick representation in light green (RRM1) and dark green (RRM2). B) A schematic representation of interactions that stabilise the orientation of the RRM1, RRM2 and two additional α-helical elements (α_N and α_L).

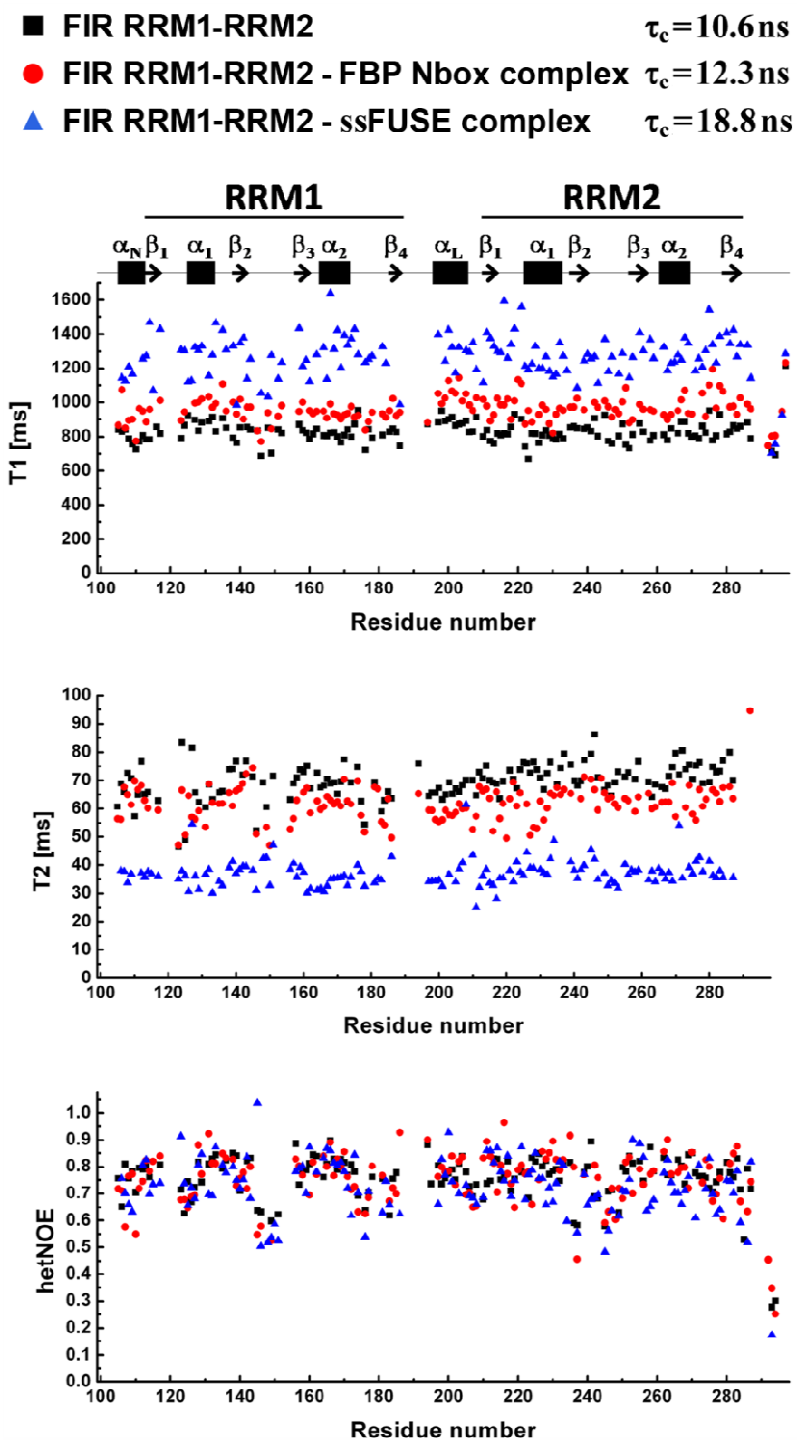

Figure 22 Relaxation data. A) Amide resonances' ^{15}N T1 (top), ^{15}N T2 (middle) and hetNOE (bottom) values for FIR RRM1-RRM2 (black squares), FIR RRM1-RRM2 – FBP Nbox (red circles) and FIR RRM1-RRM2 – ssFUSE29 (2 : 1 ratio – blue triangles) *vs.* RRM1-RRM2 sequence.

Chapter 4

Results – Structure of the FIR RRM1-RRM2 – FBP Nbox complex

4.1 FBP Nbox peptide expression and purification

The most cost- and time-efficient method of the production of isotopically enriched proteins relies on overexpression of the desired constructs in *E. coli* cells. The bacterial cells are cultured in minimal media, where the only sources of nitrogen and carbon contain ^{15}N (e.g. $(^{15}NH_4)_2SO_4$) and ^{13}C (e.g. $^{13}C_6$-D-glucose) isotopes. The recombinant proteins produced in *E. coli* cells are exposed to non-native folding conditions and/or proteases present in the bacterial cells. This is a great concern when separate domains, large unfolded regions or short peptides are to be produced.

The identified region of FBP involved in the interaction with FIR RRM1-RRM2 (FBP Nbox) is only 26 amino acids long (Chung et al., 2006). To increase the yield of the FBP Nbox peptide in *E. coli* cells it was produced as a HisTag-GST fusion. The overexpressed recombinant protein was present in a soluble fraction of lysed bacterial cells (Figure 23). The standard lysis, IMAC purification and TEV digestion protocol, used in our laboratory, allowed me to obtain an expected product of the molecular weight of ~ 3 kDa, without any signs of lower molecular weight degradation products (Figure 23: Flow-through 2 lanes). The recombinant peptide was further purified by high performance liquid chromatography (HPLC) to eliminate any traces of proteolytic activity. The final yield of the $^{15}N^{13}C$-labelled FBP Nbox peptide was ~ 0.5 mg per litre of culture.

4.2 Biophysical characterisation and assignment of the FBP Nbox peptide

Preliminary experiments on the $^{15}N^{13}C$-labelled FBP Nbox peptide at pH 8.0 showed that resonances in ^{1}H-^{15}N (but not in ^{1}H-^{13}C) correlation spectra are very broad and mostly undetectable (data not shown). This broadening is temperature- and pH-dependent (Figure 24A), indicating that it is caused by the chemical exchange of protons between amide groups of the peptide backbone and water molecules (Dempsey, 2001; Hernandez and LeMaster, 2009). Decrease in both temperature and pH lead to an increased number of detectable amide resonances and at pH 5.0 nearly all of

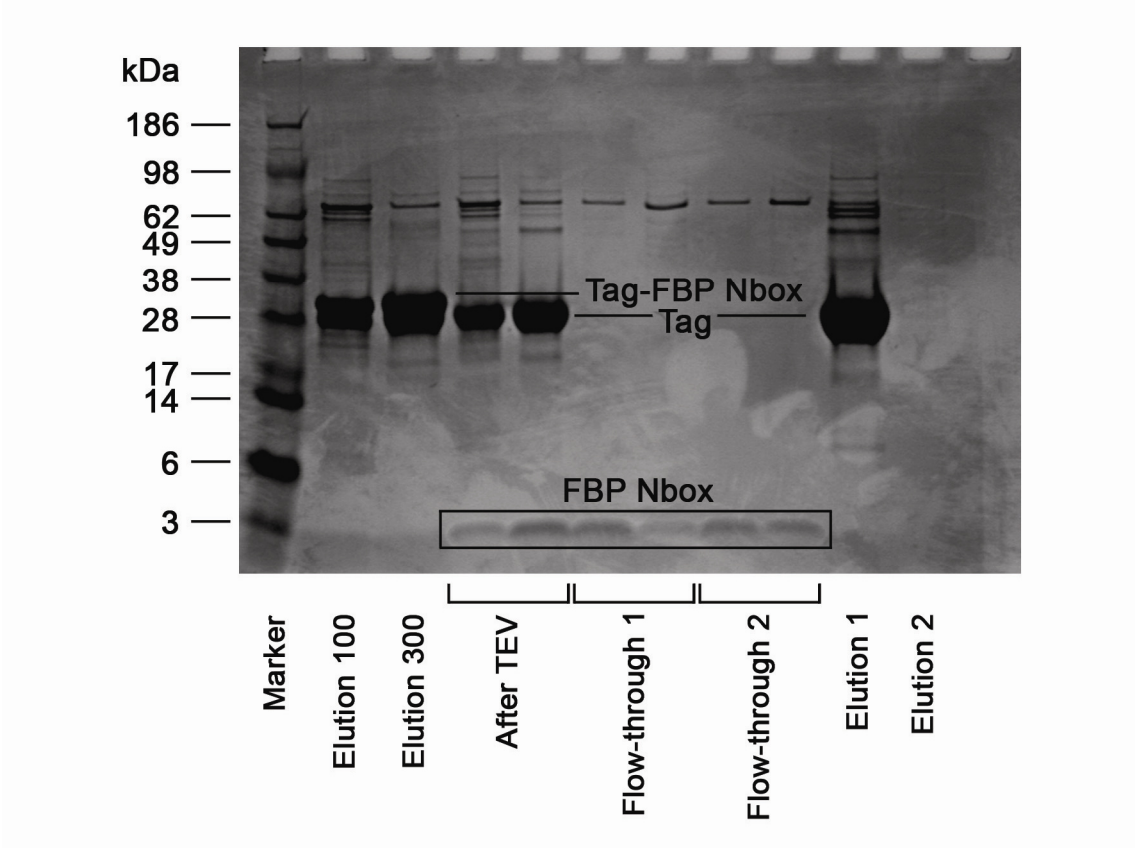

Figure 23 SDS-PAGE analysis of the FBP Nbox peptide purification. The HisTag-GST-FBP Nbox fusion protein was purified on the IMAC column from the soluble fraction of the bacterial cell lysate (Elution 100 and 300). The eluted protein was treated overnight with TEV protease (After TEV) and the fusion (HisTag-GST) was removed by a two-step purification on the IMAC column (flow-through and elution fractions 1 and 2). The recombinant FBP Nbox peptide was found in the flow-through fractions (1 and 2), while the tags retained on the columns (Elution 1 and 2).

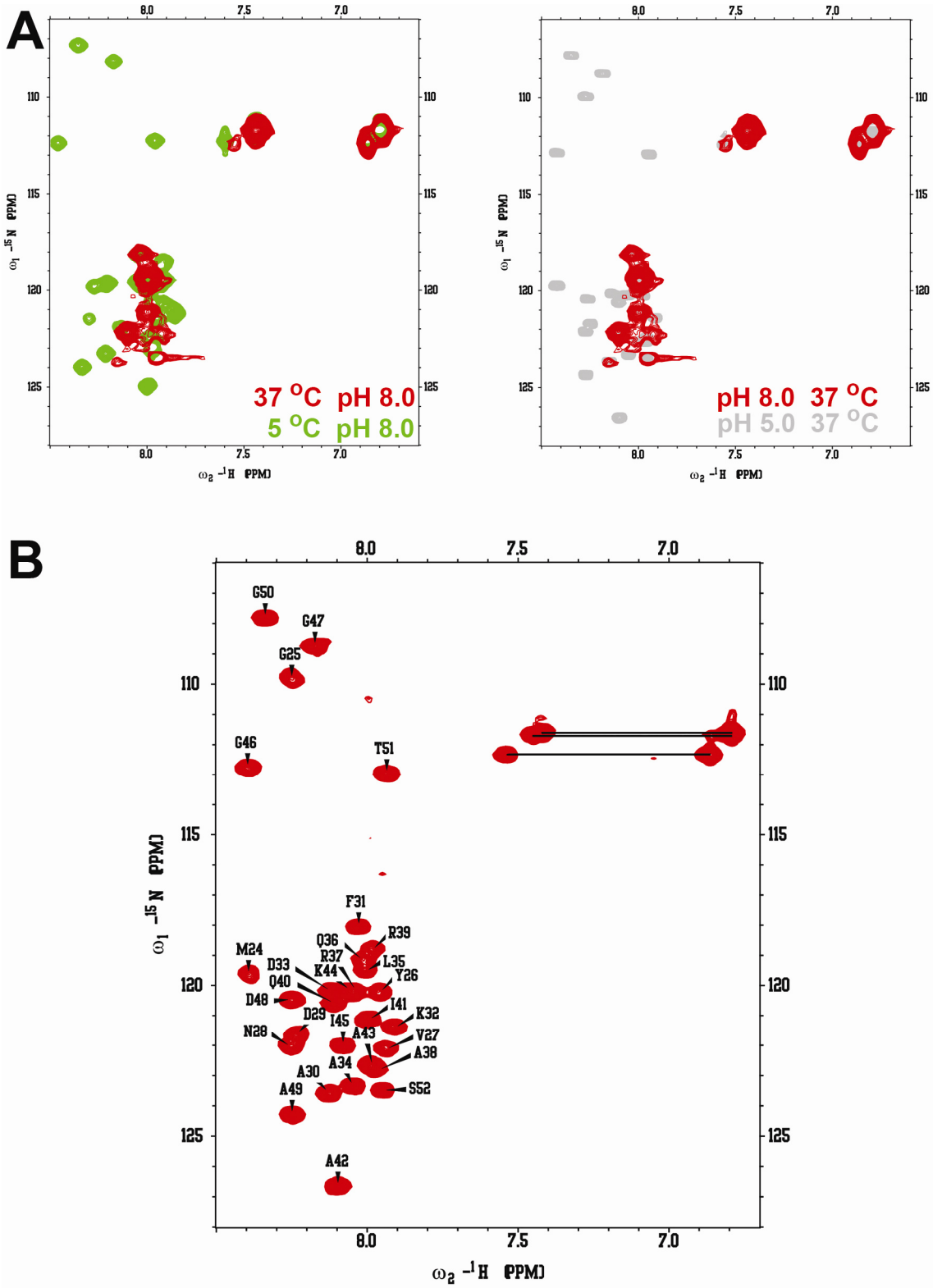

Figure 24 ^{15}N sofast HMQC spectra of the FBP Nbox. A) Experiments recorded under different conditions. The exchange of amide protons with water molecules can be reduced by decreasing either the temperature (5 °C) or pH (pH = 5.0). B) The assigned spectrum of the FBP Nbox peptide.

them (29 out of 31) appeared in the ^{15}N HMQC spectrum of the FBP Nbox peptide (Figure 24B). As expected, there were also three pairs of side-chain amide resonances of glutamines (Q36 and Q40) and asparagine (N28). pH titration (from 5.0 to 8.0) did not reveal significant CSP in ^{13}C HSQC spectra, except a few peaks that turned out to correspond to the C-terminus of the peptide (data not shown). Therefore, the assignment of the peptide was achieved at pH 5.0 where standard 3D backbone experiments (correlating H^N protons to other backbone atoms) could be recorded with sufficient sensitivity. The side-chain assignment was obtained from ^{15}N TOCSY and HCCH TOCSY through-bond experiments. The chemical shifts of side-chain resonances (e.g. C^γ, C^δ, H^γ, H^δ) of the FBP Nbox peptide have nearly random-coil values (Wishart et al., 1995), as expected for the isolated peptide. However, the chemical shifts of C^α, C^β and H^α significantly deviate from the random-coil values, suggesting that the peptide possesses some secondary structure (Figure 25A). Indeed, TALOS-derived φ and ψ torsion angles of the residues D29-I41 have values characteristic of α-helix. Furthermore, the far-UV CD spectrum shows a significant content (~ 40%) of α-helix in the solution of FBP Nbox peptide (Figure 25B).

4.3 The interaction between FIR RRM1-RRM2 and FBP Nbox peptide

Titration of the ^{15}N-labelled FIR RRM1-RRM2 with an unlabelled FBP Nbox peptide mapped the interaction surface to the solvent-exposed face of the RRM2 (area of ~ 15 x 25 Å), including the β_1/α_1, α_2/β_4 loops and α_1, α_2 helices (Figure 26A). Furthermore, CSPs limited to this well defined region indicate that the protein does not undergo global rearrangements upon peptide binding. In the reverse titrations at pH 6.5 and pH 5.0 ($^{15}N^{13}C$-labelled FBP Nbox peptide titrated with an unlabelled FIR RRM1-RRM2), 19 backbone amide resonances of the peptide were perturbed (Figure 26B). A 19-residue linear peptide is too long to be accommodated on the mapped interaction surface of FIR RRM2. However, the large number of affected resonances could be explained if the peptide interacts with the FIR RRM1-RRM2 in an α-helical conformation, and the observed CSPs are the result of not only binding, but also

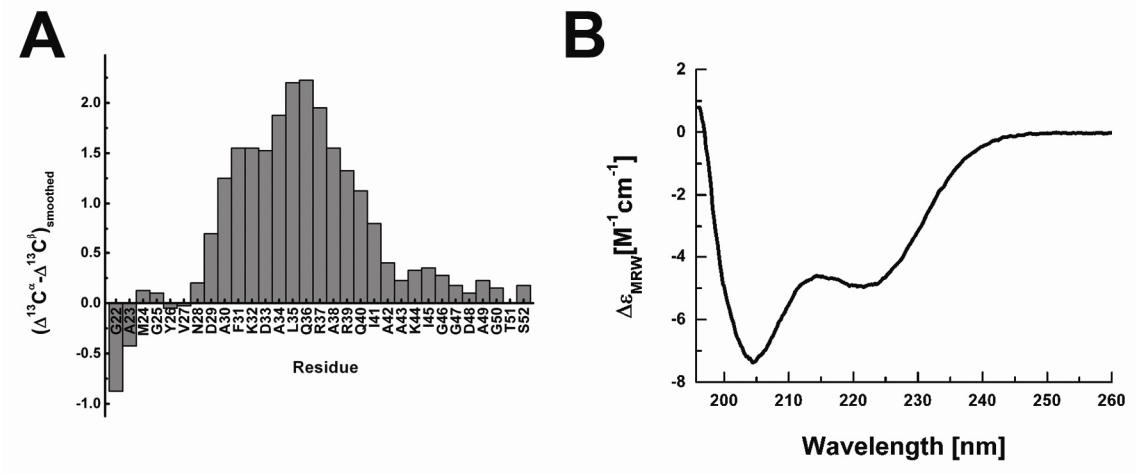

Figure 25 FBP Nbox peptide is α-helical in solution. A) Chemical shift index for the FBP Nbox peptide. The positive and negative $(\Delta^{13}C^{\alpha} - \Delta^{13}C^{\beta})_{smoothed}$ values indicate α-helices and β-strands, respectively (Metzler et al., 1993; Wishart et al., 1995). B) Far-UV CD spectrum of the free FBP Nbox peptide.

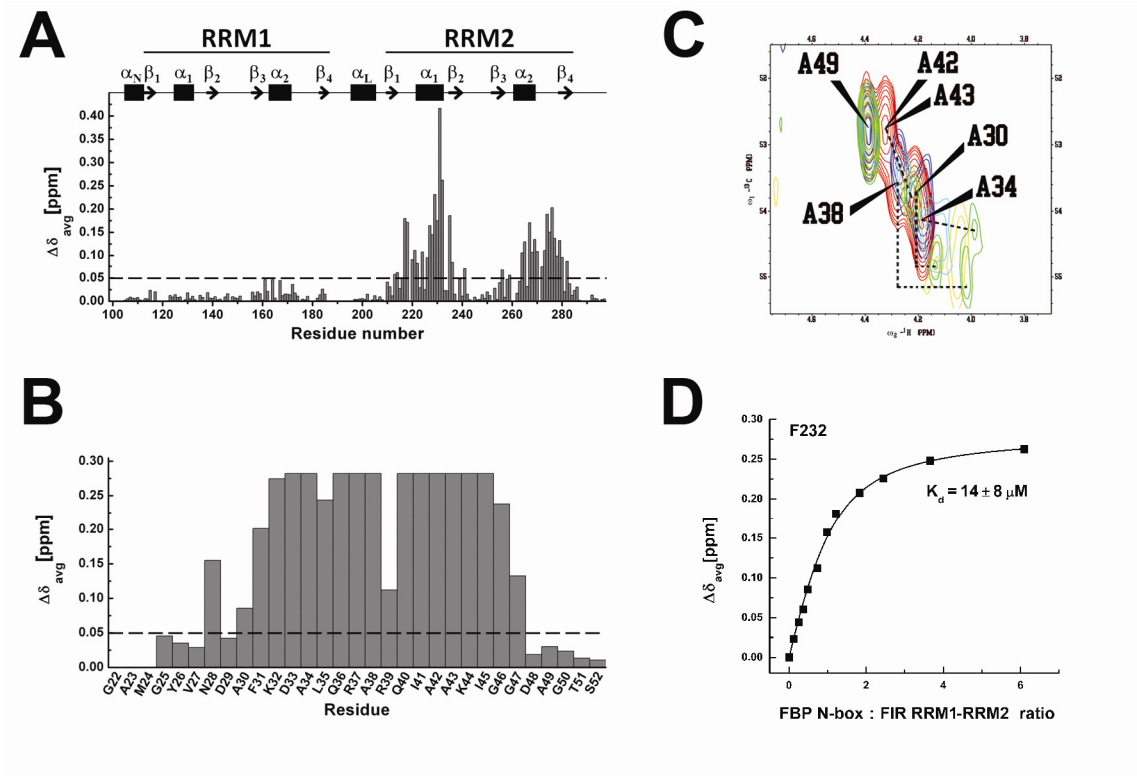

Figure 26 α-helical FBP Nbox peptide binds with low micromolar affinity to the RRM2 domain of FIR. A) Weighted chemical shift changes of the FIR RRM1-RRM2 upon addition of the FBP Nbox *vs.* protein sequence. The secondary structure elements and the boundaries of the domains are indicated above. B) Weighted chemical shift changes of the FBP Nbox upon addition of FIR RRM1-RRM2 *vs.* peptide sequence. ^{1}H-^{15}N cross-peaks, that disappear during the titration, are assigned a $\Delta\delta_{avg}$ value of 0.28. C) A section of superimposed ^{13}C HSQC spectra recorded during the titration of FBP Nbox peptide with FIR RRM1-RRM2. The spectra are colour-coded according to peptide : protein ratio: 1 : 0 – red, 1 : 0.5 – blue, 1 : 1 – cyan, 1 : 2 – yellow and 1 : 3 – green. The start and end positions of $C^{\alpha}H^{\alpha}$ cross-peaks are connected with dashed lines. The consistent downfield shifts in a ^{13}C dimension and upfield shifts in a ^{1}H dimension indicate stabilisation of the FBP Nbox peptide in an α-helical conformation upon binding to the FIR RRM1-RRM2. D) Isotherm of the FBP Nbox binding to FIR RRM1-RRM2 (NMR).

stabilisation of the secondary structure. Interestingly, $C^{\alpha}H^{\alpha}$ cross-peaks in ^{13}C HSQC spectra were consistently shifted downfield in a ^{13}C dimension and upfield in ^{1}H dimension during the titration of the $^{15}N^{13}C$-labelled FBP Nbox peptide with an unlabelled FIR RRM1-RRM2, so that the $^{13}C^{\alpha}$ and $^{1}H^{\alpha}$ chemical shift values moved towards the ones of an α-helix (Figure 26C). TALOS analysis of the chemical shifts of the peptide resonances showed that FIR binding extends the α-helical region from residues D29-I41 to V27-I45.

The perturbed resonances are in the fast-to-intermediate exchange regime during the course of the titrations and hence the shifts in the peaks position report on a fraction of the bound labelled (i.e. observed) component and can be used to obtain the K_d of the complex. The weighted CSP were plotted *vs.* the ratio between the binding partners concentrations and then fitted to a one-site binding model. The obtained K_d values are 15 ± 8 μM ($^{15}N^{13}C$-labelled FBP Nbox) and 14 ± 8 μM (^{15}N-labelled FIR RRM1-RRM2) (Figure 26D).

4.4 Structure determination of the FIR RRM1-RRM2 – FBP Nbox peptide complex

Structure determination of the FIR RRM1-RRM2 – FBP Nbox complex was achieved with the standard procedure and automatic structure calculations, however an optimisation of the experimental conditions and careful analysis were required. This is because high concentrations of FIR RRM1-RRM2 were unachievable and the complex formed a 25 kDa single unit with a dynamic interface, that lead to the broadening of the resonances. The relaxation data recorded on the complex indicate that T2 values of the FIR RRM1-RRM2 backbone ^{15}N resonances at the contact interface are shorter than expected when compared with the rest of residues (Figure 27). (The overall reduction in the T2 values in the complex is anticipated due to the higher molecular weight of the complex). To reduce the exchange broadening, and therefore improve the quality of the spectra, the subsequent experiments were recorded at different protein : peptide ratio, so the labelled component was always saturated in > 95%.

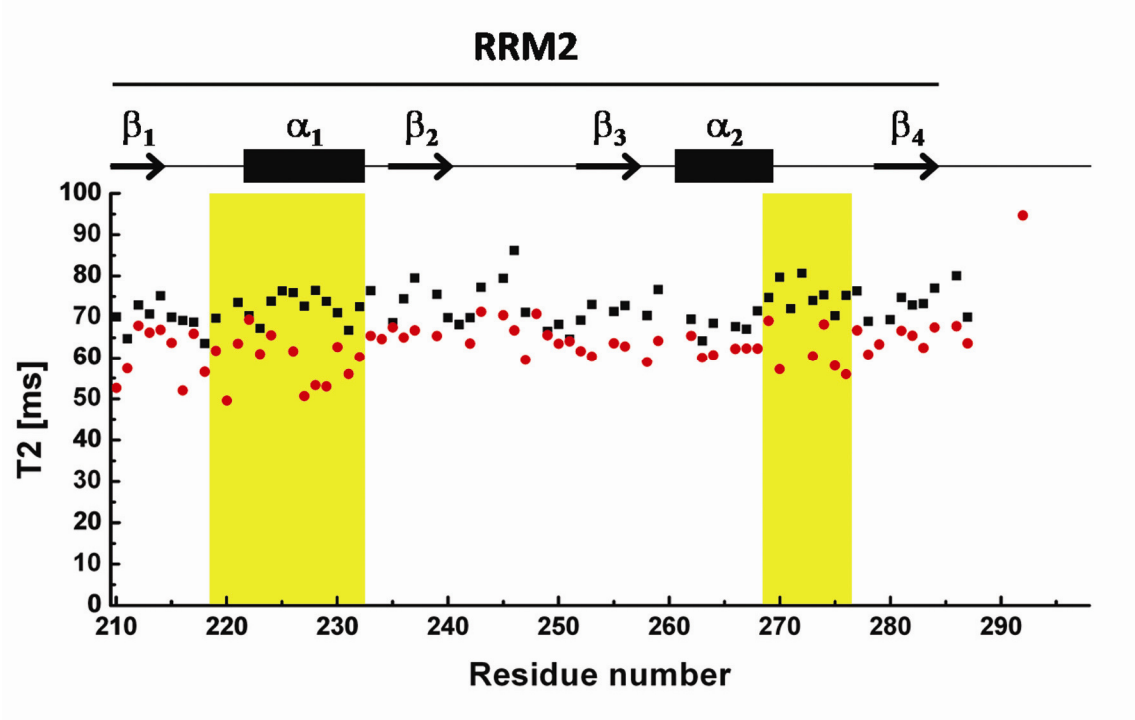

Figure 27 Amide resonances' ^{15}N T2 values of RRM2 in FIR RRM1-RRM2 (black squares) and FIR RRM1-RRM2 – FBP Nbox complex (red circles) *vs.* protein sequence. The residues contacting with the peptide (yellow shading) display lower-than-average ^{15}N T2 in the complex, but not in the free form, consistently with a dynamic interface.

4.4.1 Backbone and side-chain assignment of the complex

The 3D ^{13}C NOESY experiment recorded on $^{15}N^{13}C$-labelled FIR RRM1-RRM2 – FBP Nbox peptide complex showed that the pattern of NOE cross-peaks is very similar to that obtained for the FIR RRM1-RRM2 alone (data not shown). This means that there is no global rearrangement in the protein structure upon binding to FBP Nbox peptide, in agreement with titration experiments where the CSP is limited to one side of the protein. Because of lack of global changes and the fast-to-intermediate exchange regime the assignment of FIR RRM1-RRM2 could be transferred from the free form to the bound one in a two-step procedure. First, the assignment of backbone amide resonances was achieved following the evolution of the ^{1}H-^{15}N cross-peaks during the titration of the protein with the peptide. Then, combination of 3D ^{15}N and ^{13}C NOESY spectra of the complex allowed me to complete the assignment of the ^{13}C and ^{13}C-attached proton resonances (see Appendix V). The whole process was validated with 3D HNCA backbone experiment recorded on the complex to resolve any discrepancies due to the overlap of ^{15}N HMQC spectra from the titration.

The procedure applied for the assignment of FIR RRM1-RRM2 in the complex could have not been used for the assignment of FBP Nbox due to the exchange broadening of peptide backbone amide resonances at pH 8.0, that was still significant after complex formation. We could not reduce this exchange process by decreasing pH or temperature (as done for the free peptide), because of instability of the protein. However, there was no significant overlap in ^{13}C HSQC spectra recorded during the titration of $^{15}N^{13}C$-labelled peptide with FIR RRM1-RRM2, which allowed me to follow the evolution of ^{1}H-^{13}C cross-peaks and to transfer the assignment of ^{13}C and ^{1}H from the free form of the peptide to the bound one (see Appendix V). This assignment was cross-checked using 3D ^{13}C NOESY and 3D HCCH TOCSY experiments recorded on the complex ($^{15}N^{13}C$-labelled FBP Nbox).

4.4.2 Obtaining structural restraints

The intra-molecular distance (NOE-derived and H-bond) and angle restraints for the peptide and the protein were achieved as described for the FIR RRM1-RRM2, with one exception. Despite strong evidence of α-helical conformation of the peptide in the complex (pattern of the NOE and TALOS-derived dihedral angles) the backbone amide resonances were still in fast exchange with the water molecules, making exchange-protection experiment unfeasible. Therefore, H-bond restraints were assumed for the α-helical region of the peptide (A30-K44), identified with chemical shift values of the backbone and C^{β} resonances and with NOE patterns.

The inter-molecular restraints are crucial for the structure determination of the complex. To identify sparse inter-molecular cross-peaks among many intra-molecular ones in the ^{13}C NOESY spectra two different experiments were recorded: ^{13}C-filtered ^{13}C NOESY ($^{15}N^{13}C$-labelled FIR RRM1-RRM2) and ^{13}C NOESY without ^{13}C decoupling during the proton indirectly acquired dimension ($^{15}N^{13}C$-labelled FIR RRM1-RRM2 or $^{15}N^{13}C$-labelled FBP Nbox). In the former experiment, the initial magnetisation of ^{13}C- (protein) but not ^{12}C-attached (peptide) protons is dephased by additional pulses prior to the standard ^{13}C NOESY pulse sequence (Nietlispach et al., 2004). Consequently, NOE cross-peaks are observed only when the NOE effect occurs between ^{12}C- (peptide) and ^{13}C-attached (protein) protons. Unfortunately, the longer pulse sequence worsens the S/N in the spectra. In the latter experiment, a standard ^{13}C NOESY pulse sequence is applied, but without ^{13}C decoupling during labelling with the indirect ^{1}H frequencies. Therefore in the recorded spectra correlations between two ^{13}C-attached protons (intra-molecular) appear as a doublet, while correlations between ^{13}C- and ^{12}C-attached protons (inter-molecular) give rise to a singlet signals. Unfortunately, most of the cross-peaks in the spectrum are intra-molecular that significantly increases the spectra overlap and hampers identification of the inter-molecular correlations.

A combination of the information from the two experiments allowed me to identify 72 inter-molecular cross-peaks (Appendix VI). The assignment of ^{13}C and ^{13}C-attached protons for these cross-peaks was already known (from intra-molecular assignment). However, the assignment of the indirect ^{1}H was more complex because the unlabelled component was only ~ 35-50% bound in the sample. Therefore,

the chemical shift of a given resonance was a linear combination of the chemical shifts observed in the free and bound forms. An initial set of 14 unambiguously assigned inter-molecular cross-peaks (Table 6) from groups spanning the length of the peptide was obtained by analysis of the ^{13}C NOESY spectra (for examples see Figures 28 and 29) and used, together with intra-molecular restraints, to generate a preliminary structure of the complex. Further assignment of the inter-molecular cross-peaks was obtained iteratively, based on the already available structure.

4.4.3 Structure calculations

Structure calculations of the complex with the standard protocol used for FIR RRM1-RRM2 alone were plagued by a poor convergence, despite having identified a set of 72 inter-molecular restraints. A possible explanation for this result is the overestimation of the inter-molecular distances by the automatic calibration procedure of ARIA, that does not take into account the general broadening of resonances at the interface. To compensate the different relaxation properties of the resonances at the interface a conservative approach was taken, where all inter-molecular distances were calibrated semi-quantitatively and the upper bounds for all intra-molecular distance restraints were set up to 6 Å. This procedure relies on a high redundancy of the set of FIR RRM1-RRM2 intra-molecular restraints, that counterweight the loss of precision in the defined distances. The obtained family of structures is converged and well-defined with RMSD from the mean structure of 1.00 Å for backbone atoms and 1.45 Å for all heavy atoms (Table 7 and Figure 30).

4.5 Description of the structure of FIR RRM1-RRM2 – FBP Nbox complex

The structure of FIR RRM1-RRM2 in the complex with FBP Nbox is very similar to the structure of the protein alone (Figure 31A and see Figure 20C) with

Table 6 A set of 14 inter-molecular restraints used at the initial stage of structure calculations of FIR RRM1-RRM2 – FBP Nbox complex.

No.	FIR RRM1-RRM2 atom(s)	FBP Nbox peptide atom(s)	Upper bound [Å]
1	**F232 HA**	**A30 QB**	**3.5**
2	**V228 QG1**	**L35 QQD**	**3.5**
3	**A231 QB**	**A30 QB**	**5.0**
4	**A231 QB**	**D33 HB2**	**3.5**
5	**A231 QB**	**D33 HB3**	**3.5**
6	**D224 HB2**	**I41 QG2/I41 QD1**	**5.0**
7	**D224 HB3**	**I41 QG2/I41 QD1**	**5.0**
8	**V228 QG1**	**I41 QG2/I41 QD1**	**2.8**
9	**V228 QG2**	**I41 QG2/I41 QD1**	**5.0**
10	**L220 QD1**	**I41 QG2/I41 QD1**	**3.5**
11	**L220 QD2**	**I41 QG2/I41 QD1**	**3.5**
12	**L275 QD1**	**I41 QG2/I41 QD1**	**2.8**
13	**S269 HB2**	**V27 QQG**	**5.0**
14	**S269 HB3**	**V27 QQG**	**5.0**

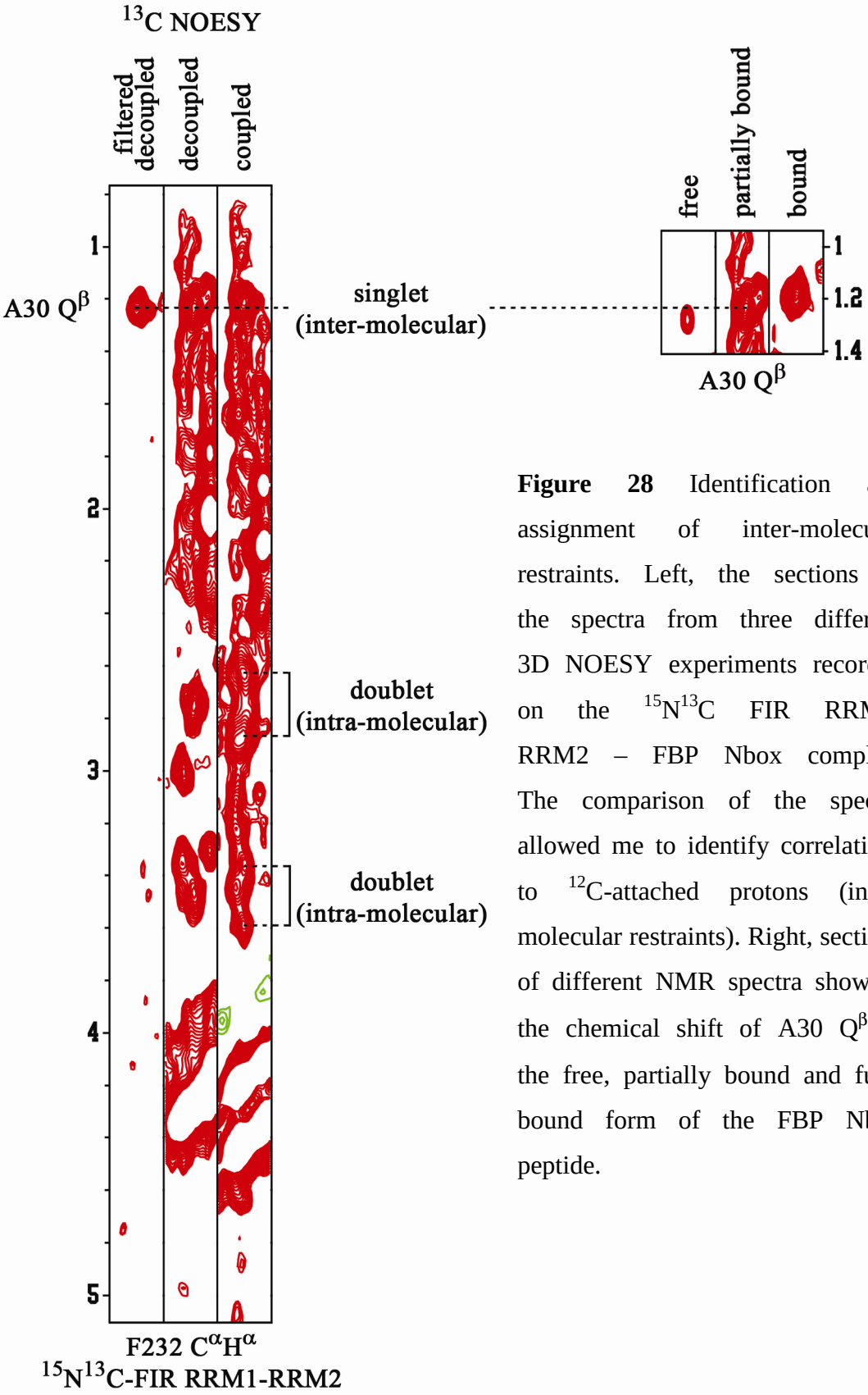

Figure 28 Identification and assignment of inter-molecular restraints. Left, the sections of the spectra from three different 3D NOESY experiments recorded on the $^{15}N^{13}C$ FIR RRM1-RRM2 – FBP Nbox complex. The comparison of the spectra allowed me to identify correlations to ^{12}C-attached protons (inter-molecular restraints). Right, sections of different NMR spectra showing the chemical shift of A30 Q^{β} in the free, partially bound and fully bound form of the FBP Nbox peptide.

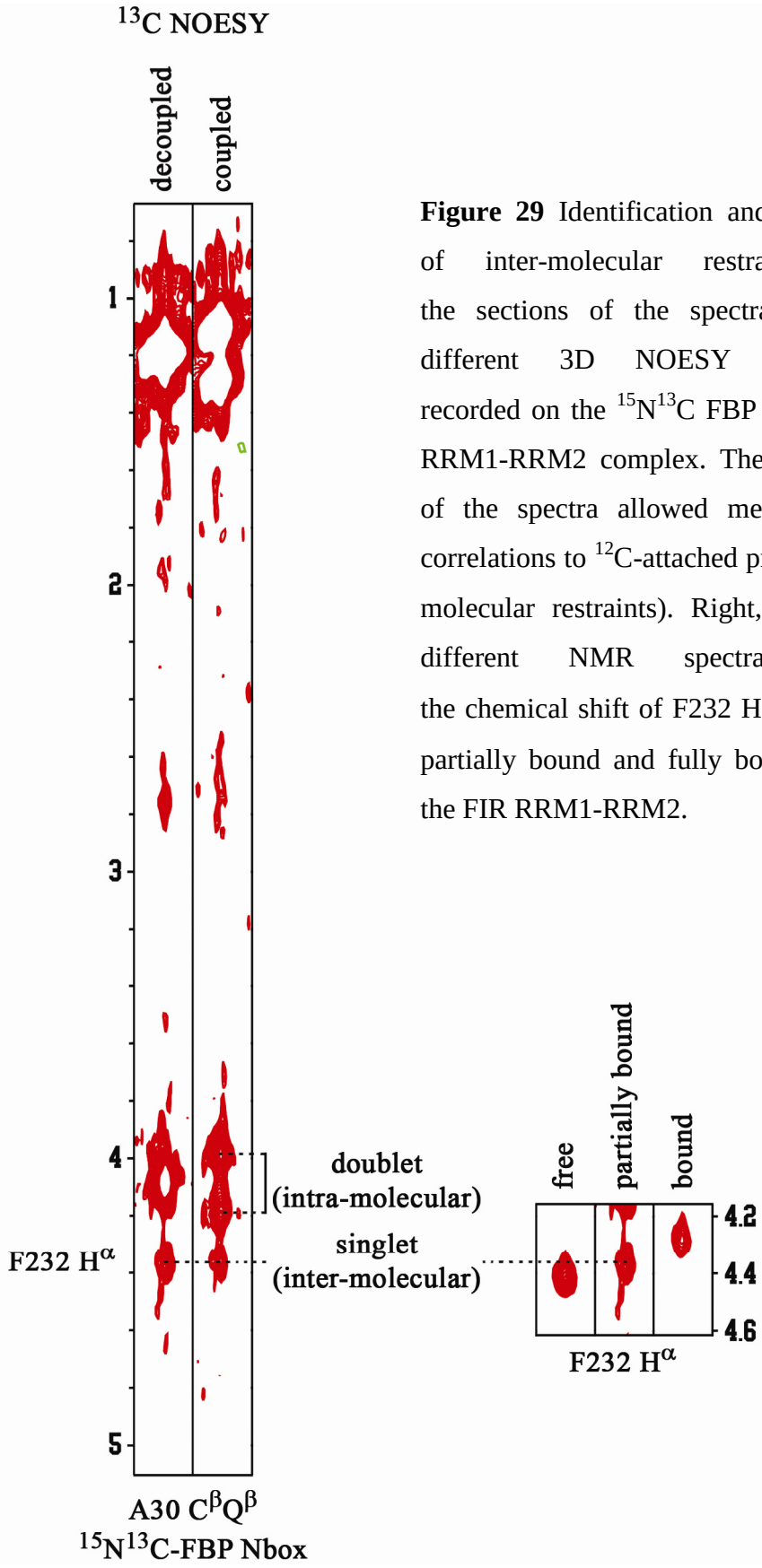

Figure 29 Identification and assignment of inter-molecular restraints. Left, the sections of the spectra from two different 3D NOESY experiments recorded on the $^{15}N^{13}C$ FBP Nbox – FIR RRM1-RRM2 complex. The comparison of the spectra allowed me to identify correlations to ^{12}C-attached protons (inter-molecular restraints). Right, sections of different NMR spectra showing the chemical shift of F232 H^{α} in the free, partially bound and fully bound form of the FIR RRM1-RRM2.

Table 7 NMR statistics for FIR RRM1-RRM2 – FBP Nbox complex.

	FIR RRM1–RRM2 (103-297) – FBP Nbox (27-52) complex
NMR distance and dihedral restraints	
Distance restraints	4611
Intra-residue	1632
Inter-residue	2979
Sequential ($\|i - j\| = 1$)	878
Medium-range ($1 < \|i - j\| \leq 4$)	546
Long-range ($\|i - j\| > 4$)	1483
Inter-molecular	72
TALOS φ and ψ restraints	280
Hydrogen bond restraints	36
Structure statistics	
Mean total energy (kcal/mol)	-7598 ± 180
Dihedral angles violations > 5º	2.2 ± 0.9
NOE violations > 0.5Å	0
NOE violations > 0.3Å	2.6 ± 1.2
Mean NOE energy (kcal/mol)	67 ± 6
RMSD from idealised covalent geometry	
Bond lengths (Å)	0.0036 ± 0.0001
Bond angles (º)	0.507 ± 0.008
RMSD from the mean structure (Å)	(res. 105-288, 27-45)
Backbone atoms	1.00 ± 0.09
Heavy atoms	1.45 ± 0.08
RMSD from the mean structure (Å)	secondary structure element only*
Backbone atoms	0.71 ± 0.08
Heavy atoms	1.18 ± 0.08
Ramachandran plot analysis	
Most favoured regions (%)	82.2
Additional allowed regions (%)	15.3
Generously allowed regions (%)	1.0
Disallowed regions (%)	1.5

* residues 105-111, 113-117, 125-132, 138-143, 155-160, 164-172, 184-185, 194-208, 210-214, 222-229, 235-243, 248-257, 260-269 and 281-284 of FIR RRM1-RRM2 and residues 30-44 of FBP Nbox

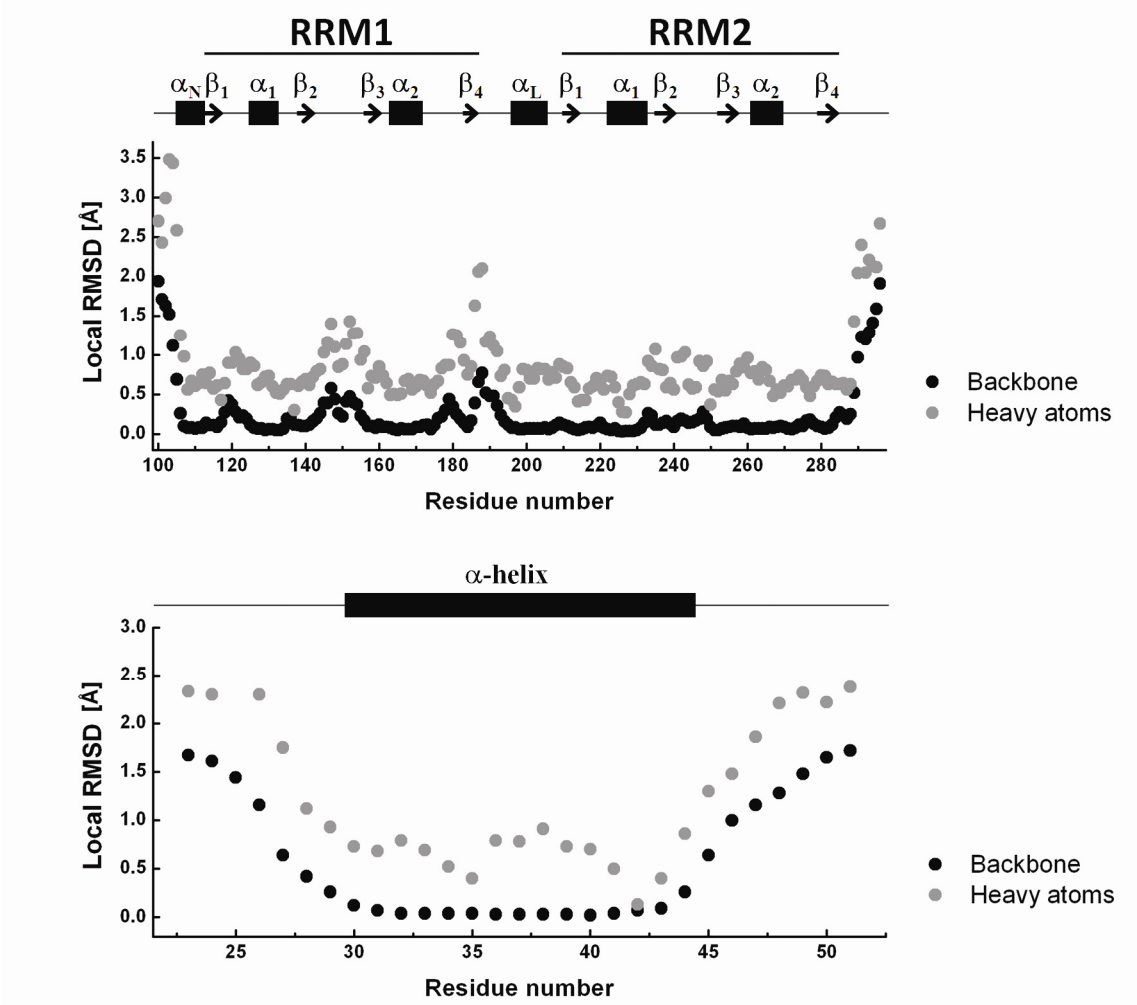

Figure 30 Local RMSD in the complex *vs.* FIR RRM1-RRM2 (top) and FBP Nbox (bottom) sequence. The 20 calculated conformers were superimposed in Molmol (Koradi et al., 1996) using residues 105-111, 113-117, 125-132, 138-143, 155-160, 164-172, 184-185, 194-208, 210-214, 222-229, 235-243, 248-257, 260-269 and 281-284 of FIR RRM1-RRM2 and residues 30-44 of FBP Nbox and the local RMSD were calculated for backbone and heavy atoms. The highest RMSD are observed at the N- and C-termini and in the loop regions.

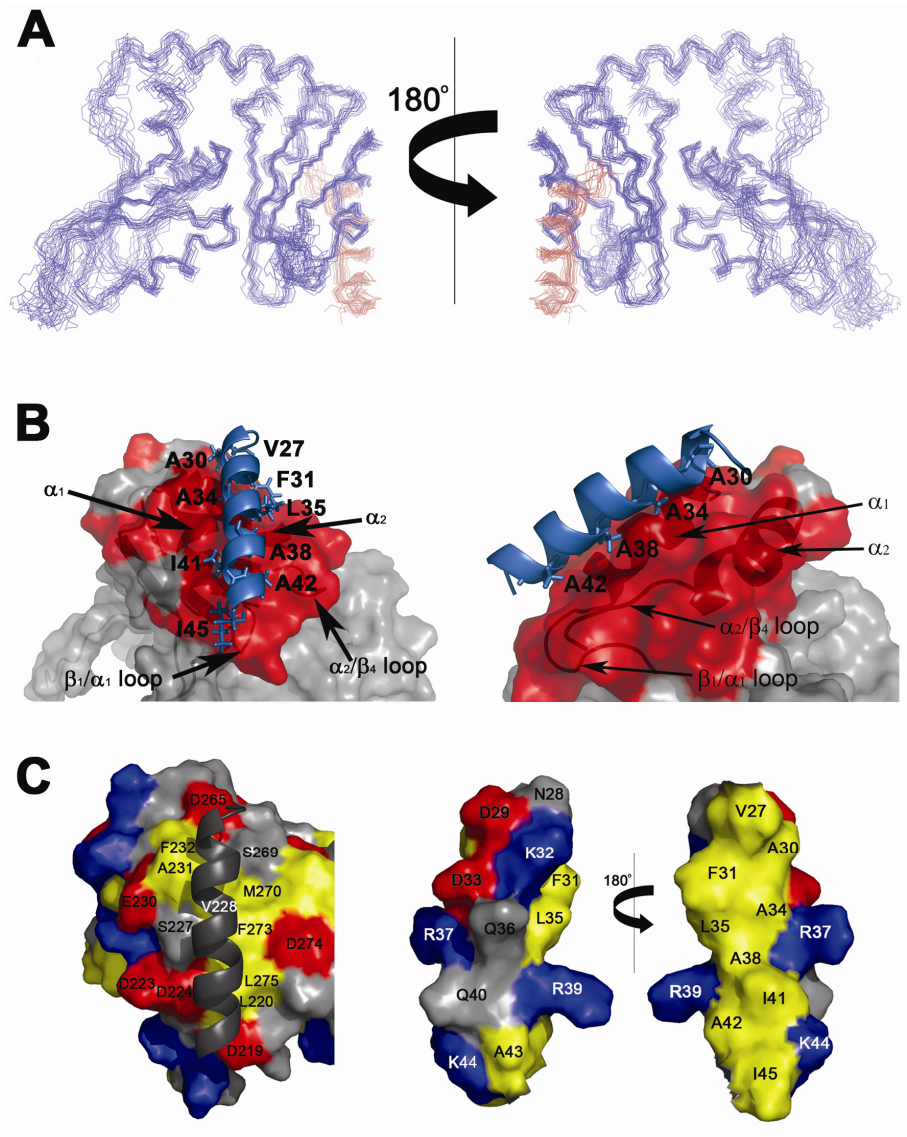

Figure 31 Hydrophobic contacts mediate the interaction between FIR RRM1-RRM2 and FBP Nbox peptide. A) Backbone trace of the 20 lowest-energy conformers of the FIR RRM1-RRM2 (blue) – FBP Nbox (brown) complex. B) FBP Nbox peptide (ribbon representation – blue) bound to FIR RRM1-RRM2 (surface representation – only RRM2 is visible). FIR RRM1-RRM2 residues showing substantial CSP upon binding to the peptide are coloured in red. The hydrophobic residues of FBP Nbox involved in the interaction (left) and the four alanine residues of FBP Nbox in the central part of interface (right) are labelled. The FIR RRM1-RRM2 secondary structure elements involved in the peptide binding are indicated. C) Surface representation of FIR RRM1-RRM2 (only RRM2 visible – left), of the solvent exposed surface of FBP Nbox (middle), and of the FIR-interacting surface of FBP Nbox (right). Hydrophobic residues are shown in yellow, residues with positive and negative charges are displayed in blue and red respectively.

RMSD of 2.0 Å for backbone atoms and 2.4 Å for all heavy atoms between averaged free and bound structures. The relaxation measurement did not reveal any large changes in the protein dynamics, with the exception of the changes observed in the inter-molecular interface, that reflects the dynamic nature of the interaction (see Figure 22). In general we observe larger T1 and smaller T2 values for the FIR RRM1-RRM2 in the complex, with a corresponding increase in the rotational correlation time (τ_c = 12.3 ns *vs.* τ_c = 10.6 ns for FIR RRM1-RRM2 alone). These changes are consistent with the larger molecular weight of the complex (24.6 kDa *vs.* 21.9 kDa for the protein alone).

The bound FBP Nbox peptide adopts α-helical conformation and is located in a shallow groove between α_1 helix and β_1/α_1 loop on one side and α_2 helix and α_2/β_4 loop on the other (Figures 31 and 32). The contact interface matches very well CSP data obtained from the titration experiments (Figure 31B). The N-terminus of the FBP Nbox is located where α_1 and α_2 helices of FIR RRM2 meet and the C-terminus is docked between β_1/α_1 and $\alpha_2/\beta4$ loops, so the peptide covers essentially the entire length of RRM2 domain. The interaction is hydrophobic (Figure 31C) and involves a narrow stretch of residues on RRM2 (F232, A231, V228, M270, F273, L275 and L220), that provide binding platform for the hydrophobic face of FBP Nbox peptide (Figure 31C and 32). Residues L220, L275 and F273 of FIR form a hydrophobic pocket that accommodates the side-chain of I41 of the FBP Nbox peptide; F273 and V228 make hydrophobic interactions with A38 and finally, A34 of the peptide is sandwiched between A231, V228, F232 (RRM2) and F31 (FBP Nbox peptide). Interestingly, four alanines are found among nine residues that form hydrophobic side of the peptide (Figure 31C). The central area of the interface is occupied by two of those alanines (A34 and A38), while the bulky hydrophobic residues are loosely packed against the protein. Such an arrangement results in a weak inter-digitation between the side-chains of the peptide and protein and ensures that the interaction is specific, but transient, as required for the biological function of FIR.

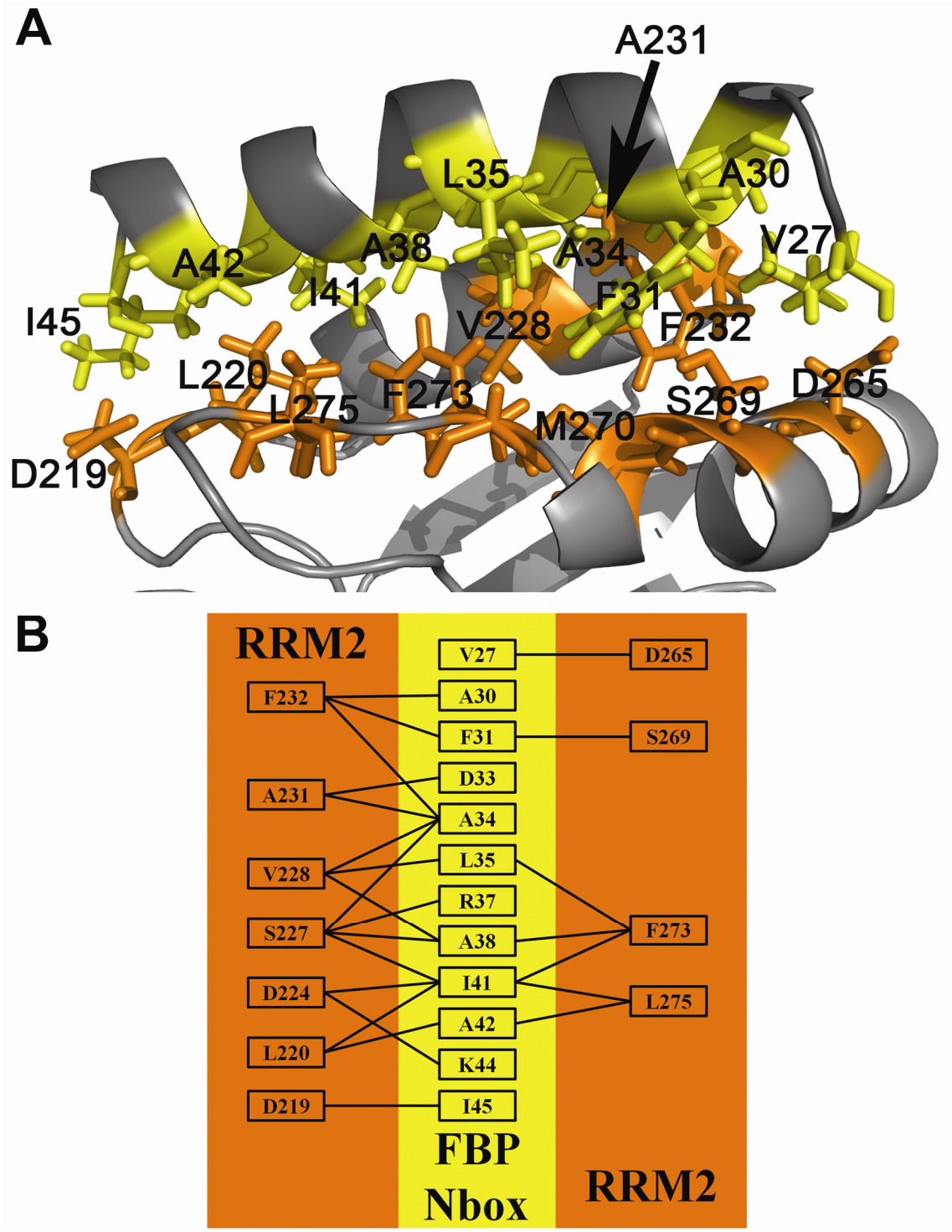

Figure 32 Detailed view of the FIR RRM2 – FBP Nbox interface. A) The peptide (yellow) and protein (orange) residues show no significant inter-digitation between side-chains. B) A schematic representation of the protein–peptide interaction in the complex.

Chapter 5

Results – Protein–protein and protein–DNA interactions in the FUSE–FBP–FIR system

5.1 Comparison of the binding of three FBPs Nbox peptides to FIR RRM1-RRM2

The FBP family consists of three members: FBP (FBP1), FBP2 and FPB3. Interestingly, in a yeast two-hybrid assay FIR RRM1-RRM2 interacts with FBP and FBP2, but not FBP3, Nbox sequences (Chung et al., 2006). I show for the first time that all three peptides directly interact with FIR RRM1-RRM2 (Figure 33). NMR titration experiments indicate that the surface of interaction is conserved across the FBP family (Figure 33B). However, the exchange regime of the affected resonances show that individual complexes have different binding kinetics. The protein resonances are in the fast-to-intermediate exchange regime on the chemical shift timescale during the titrations with FBP and FBP2 Nboxes and in the fast exchange regime during the titration with FBP3 Nbox. The difference in the exchange regime may reflect the differences in the dissociation constants of the complexes. Indeed, fitting of the binding curves to one-site binding model shows that FIR RRM1-RRM2 binds Nboxes of FBP and FBP2 with similar affinities (K_d = 14 ± 8 μM and K_d = 6 ± 6 μM respectively), while the affinity for FBP3 Nbox is ~ 30-fold weaker (K_d = 280 ± 36 μM) (Figure 33C).

The origin of differences observed in the affinity of FIR RRM1-RRM2 for Nboxes has not been addressed experimentally, but it is possible to make some hypotheses based on the available structure of the complex, peptides sequence alignment and the α-helicity of the peptides. Sequence alignment revealed that Nboxes of FBP and FBP2 proteins share 77% identity, while FBP3 Nbox is more divergent (46% and 42% identity with Nboxes of FBP and FBP2, respectively) (Figure 33A). Interestingly, the Nboxes of FBP and FBP3 differ only in three residues (V27, A30 and A38 of FBP) at the protein–peptide interface. At the first position, all three peptides have different residues and FBP3 (lysine) is more similar to FBP2 (arginine) than to FBP (valine). However, 1) this position is at the edge of the contact interface and 2) FBP3 binds weaker to FIR RRM1-RRM2 than the other two peptides and thus, this position is unlikely to define the affinity of the complexes. At the second and third different positions, Nboxes of FBP and FBP2 are identical (two alanines), while in FBP3 there are glycine and valine, respectively. Replacement of A30 (in FBP) by glycine (in FBP3) leads to the decrease in the hydrophobic contact interface and thus,

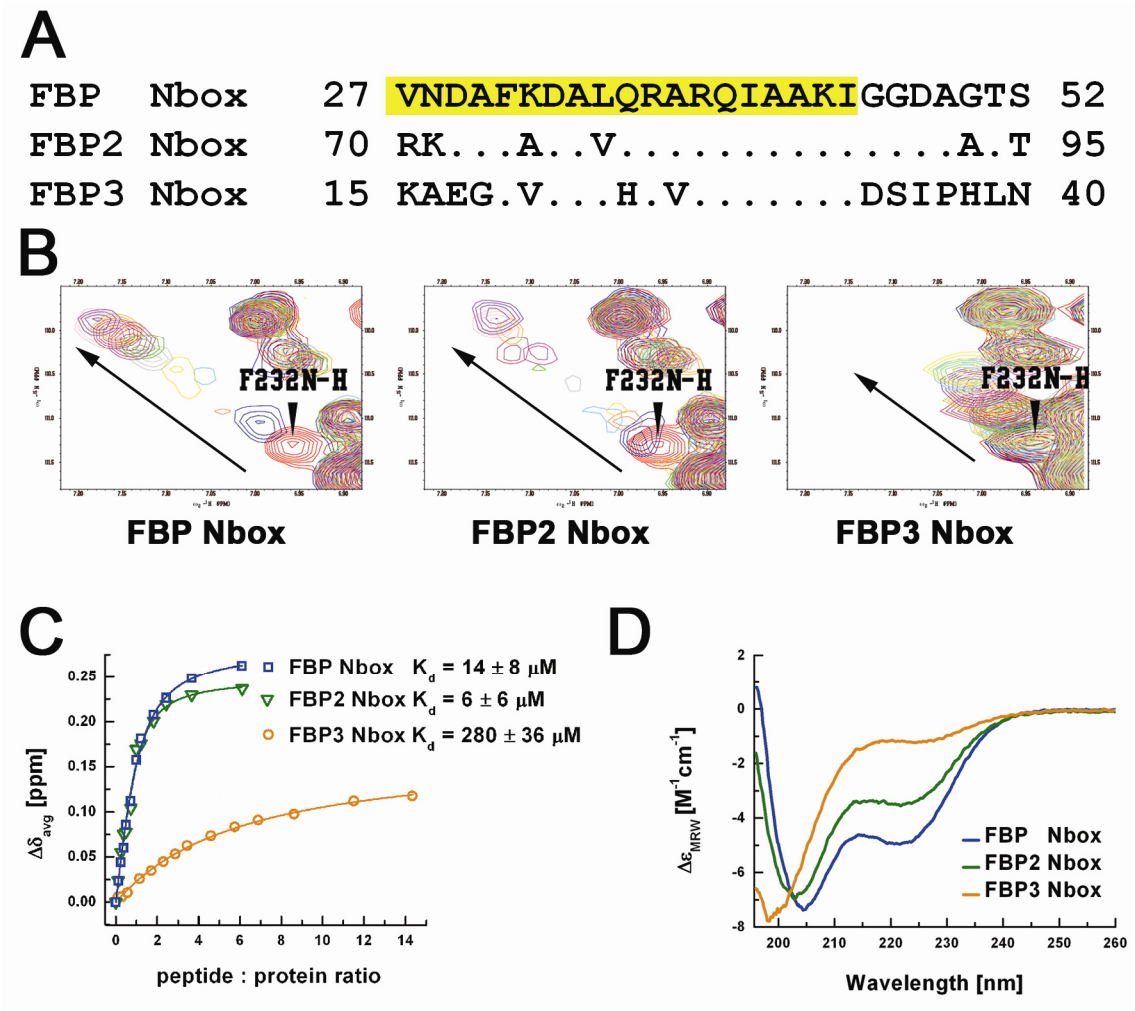

Figure 33 Comparison of the binding of three different FBPs Nbox peptides to FIR RRM1-RRM2. A) Sequence alignment of Nbox sequences from FBP, FBP2 and FBP3 proteins. The dots represent identical residues to FBP Nbox sequence. The positions of the amino acids in the protein sequence is indicated. The residues of FBP Nbox that form an α-helix upon interaction with the FIR RRM1-RRM2 are highlighted in yellow. B) Sections of superimposed ^{1}H-^{15}N correlation spectra recorded during titrations of FIR RRM1-RRM2 with FBP Nbox (left), FBP2 Nbox (middle) and FBP3 Nbox (right). The resonances are in fast-to-intermediate exchange regime during the titration with FBP and FBP2 Nboxes and in fast exchange regime during the titration with FBP3 Nbox. C) Isotherms of the binding of different Nboxes to FIR RRM1-RRM2. The affinity of the protein for FBP3 Nbox is ~ 30-fold weaker than for Nboxes of FBP and FBP2. D) Far-UV CD spectra of the FBP, FBP2 and FBP3 Nbox peptides. FBP and FBP2 Nboxes have a significant content of α-helix.

may cause the decrease in the affinity. Replacement of A38 (in FBP) by valine (in FBP3) places a large hydrophobic side-chain in the centre of the interface causing a steric hindrance. Indeed, the available mutational analysis of the FBP Nbox peptide confirms these observations (Chung et al., 2006). Apart from different amino acid sequence, FBP3 Nbox peptide has lower propensity to form an α-helix than the Nboxes of FBP and FBP2 have – 0% *vs.* ~ 40% of α-helix content as assessed by CD (Figure 33D). This may have important consequences for the affinity of the protein for the peptide, because this interaction can be described by a two-step model, where the peptide adopts α-helical conformation and then interacts with the protein:

$$\text{Peptide}_{\text{random-coil}} \rightleftarrows \text{Peptide}_{\alpha\text{-helix}} \rightleftarrows \text{Peptide} - \text{Protein complex}$$

Therefore, the weaker affinity of FIR RRM1-RRM2 for FBP3 Nbox may arise from both the substitutions at the contact interface and the α-helix propensity of the peptide.

5.2 DNA sequence preference of FIR RRM1

The scaffold-independent analysis (SIA) of FIR RRM1 described in this section has been performed with the help of Dr. Irene-Diaz Moreno at the National Institute for Medical Research, London, UK.

It has been reported that FIR RRM1-RRM2 interacts with ssFUSE DNA only via its first domain (Crichlow et al., 2008). CSP data obtained by titrating FIR RRM1-RRM2 with a ssFUSE29 DNA (Chung et al., 2006) confirmed that the DNA binds the nucleic acid recognition surface of RRM1 but not the one of RRM2, that is indeed engaged in the interaction with RRM1 (Figure 34). The DNA binding surface mapped by CSP encompasses most of the residues of the β-sheet as well as those of β2/β3 and α2/β4 loops of RRM1, including the residues that make direct contacts with the only nucleotide visible in the published X-ray structure (Crichlow et al., 2008).

Despite FIR protein being linked to transcription regulation (Liu et al., 2000) and splicing (Page-McCaw et al., 1999) little is known about its nucleic acid sequence preference. Missing-base interference footprinting revealed that FIR can bind to

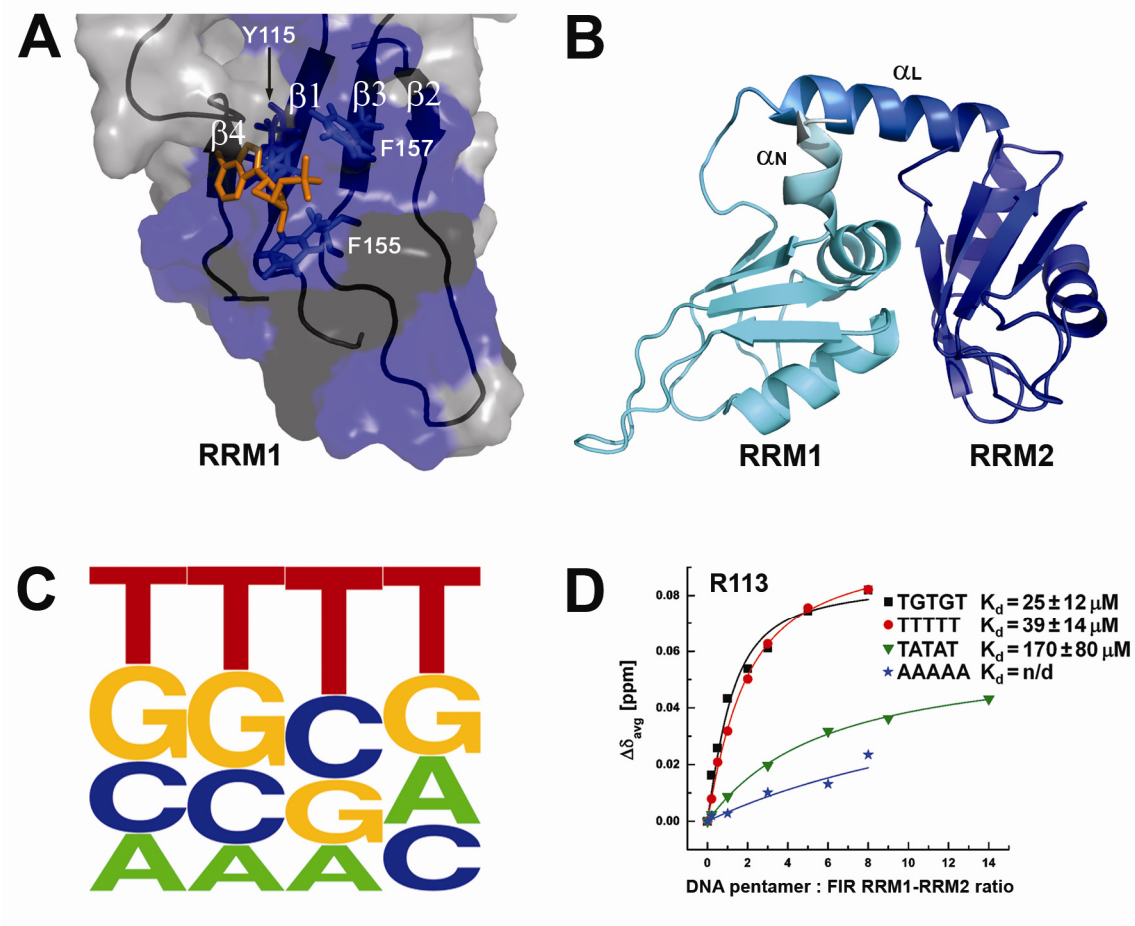

Figure 34 FIR interacts preferentially with T and G-rich sequences via its first RRM domain. A) DNA binding surface of FIR RRM1. Residues showing significant CSP upon titration with the ssFUSE29 are in blue and residues whose amide resonance are not assigned are in black. The position of the single nucleotide visible in the X-ray structure (orange) is reported on the FIR RRM1-RRM2 structure. The side-chains of the aromatic residues located in the conserved RNP2 and RNP1 nucleic acid binding motifs (Y115, F155, F157) are displayed. The DNA molecule binds the canonical nucleic acid binding surface of the RRM1 domain. B) Ribbon representation of FIR RRM1-RRM2. The β-sheet of RRM2 (dark blue), that is a putative DNA binding surface, is engaged in the inter-domain contacts with RRM1 (light blue). C) Graphic representation of FIR RRM1-RRM2 DNA sequence preference. The picture was generated by plotting the SIA data with the Weblogo server (http://weblogo. berkeley.edu/logo.cgi). D) NMR binding isotherms of four ssDNA pentamers interacting with FIR RRM1-RRM2 and the measured dissociation constants (n/d – not determined).

multiple T-rich sites on ssFUSE DNA (Benjamin et al., 2008). However, the only visible nucleotides in the X-ray structure of FIR RRM1-RRM2 dimer bound to ssFUSE DNA are adenine and cytosine (Crichlow et al., 2008). Typically, RRM domains recognise four nucleotides (Maris et al., 2005) and we used SIA (Beuth et al., 2007) to characterise ssDNA sequence preference of FIR RRM1 in four nucleotide positions. The results showed that the protein prefers Ts and Gs and selects against As in all measured positions, returning a consensus sequence of (T/G)(T/G)TT (Figure 34C and Table 8).

The findings were validated by testing the binding of four *ad hoc* 5mer ssDNAs (TTTTT, TGTGT, TATAT, AAAAA) to FIR RRM1-RRM2. The T and TG-rich oligonucleotides, which are in fact present in the FUSE sequence, bind the protein with the same affinity (K_d = 39 ± 14 μM and K_d = 25 ± 12 μM respectively). Inclusion of two adenines into the sequence (TATAT) leads to a five-fold drop in affinity (K_d = 170 ± 80 μM), while a polyA sequence does not show significant binding (Figure 34D).

5.3 Interaction between FIR RRM1-RRM2 and ssFUSE

The BLI experiments on the FIR RRM1-RRM2 – ssFUSE40 DNA interaction described in this section have been performed with the help of Dr. Stephen R. Martin at the National Institute for Medical Research, London, UK.

The moderate sequence preference of FIR RRM1 suggests that the protein will bind to multiple sites within the longer T-rich FUSE sequence. Indeed, NMR titration of FIR RRM1-RRM2 with the ssFUSE29 showed that the protein is saturated at ~ 0.4-0.6 DNA : protein ratio suggesting at least 2 : 1 stoichiometry of the binding. The resonances perturbed during the titration are in a fast-to-intermediate exchange regime, indicating that the K_d is in low micromolar range. However, the conditions of the experiment and the quality of data did not allow me to obtain accurate K_d. The FIR – DNA interaction was further explored with BLI using a FUSE-derived 40mer ssDNA – a sequence that encompasses ssFUSE29 (see Table 4). The longer DNA includes all the nucleotides that are protected by FBP and FIR binding in missing-base interference

Table 8 SIA scores for FIR RRM1-RRM2. 15 peaks were analysed for each DNA pool.

FIR	NXNNN	NNXNN	NNNXN	NNNNX
A	**0.49**	**0.42**	**0.34**	**0.61**
C	**0.62**	**0.65**	**0.61**	**0.60**
G	**0.81**	**0.84**	**0.49**	**0.74**
T	**0.86**	**0.89**	**0.93**	**0.97**
	15 peaks	**15 peaks**	**15 peaks**	**15 peaks**

experiments (Benjamin et al., 2008). This sequence was used because our primary aim was to characterise tri-molecular interactions in the system. Biotinylated ssFUSE40 was immobilised on streptavidin-coated biosensors. The association kinetics was studied by exposing sensors to different concentration of FIR (0.31 to 10 μM). After completion of binding the dissociation kinetics was monitored by moving FIR-loaded sensors to wells containing buffer (Figure 35A, left). For simplicity, the analysis was performed with the assumption that the multiple binding sites for FIR are equivalent and non-interacting. Kinetic curves were analysed using a double exponential fit to account for a minor component probably associated with some residual non-specific binding. The observed rate constant for the major part of the association phase (k_{obs}) was linearly dependent on FIR concentration and analysis yielded association and dissociation rate constants of $9.1 \pm 0.6 \times 10^4$ $M^{-1}s^{-1}$ and 0.68 ± 0.03 s^{-1} respectively ($K_d = 7 \pm 1$ μM) (Figure 35A, right). Examination of the dissociation phase yielded an independent measure of the dissociation rate constant (0.74 ± 0.05 s^{-1}) and analysis of the dependence of the amplitude of the BLI signal on FIR concentration yielded an independent estimate of the dissociation constant (7 ± 2 μM).

The crystal structure of FIR RRM1-RRM2 in the complex with ssFUSE shows that the protein forms a dimer in the asymmetric unit (Crichlow et al., 2008). Based on this and on a 2 : 1 stoichiometry for the complex obtained from SEC-LS/RI/UV (size exclusion chromatography coupled with light scattering, refractive index and ultraviolet absorbance) data it has been proposed that FIR RRM1-RRM2 dimerises on the ssFUSE DNA (Crichlow et al., 2008). We explored this possibility by performing NMR titration experiments. As already mentioned, the DNA interacts with FIR RRM1-RRM2 with at least 2 : 1 stoichiometry, consistently with the results of Crichlow *et al.* (Crichlow et al., 2008) (Figure 35B). The observed CSP were analysed with respect to any shift at the potential dimer interface (Figure 35C). There are 11 amide protons within 10 Å of the interface in the FIR RRM2, for seven of them the resonances are assigned and non-overlapping. Interestingly, none of these resonances is affected upon the binding of ssFUSE29, suggesting that the dimer is not formed. In RRM1, the proposed dimer interface is in the proximity to the DNA binding surface and therefore it is likely to observe CSP for residues in the putative dimer interface. To decouple dimerisation from DNA binding, I compared the CSP obtained during titration with the ssFUSE29 and a DNA 5mer that binds with a 1 : 1 stoichiometry (Figure 35D). The patterns of CSP are

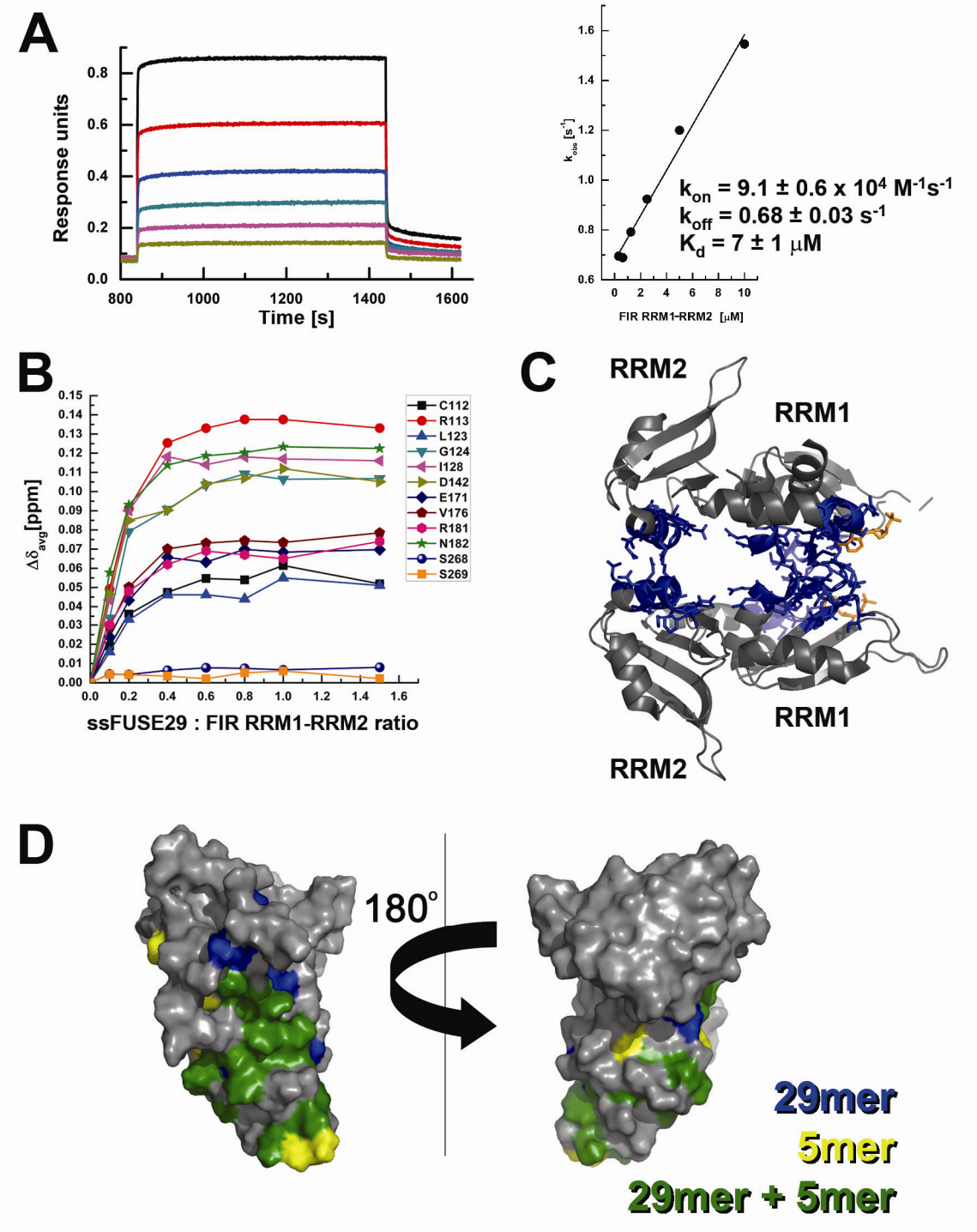

Figure 35 DNA binding and protein dimerisation by FIR RRM1-RRM2. A) Binding of increasing concentrations of FIR RRM1-RRM2 (0.31 μM, 0.63 μM, 1.25 μM, 2.5 μM, 5 μM and 10 μM) to ssFUSE 40 (left) and the corresponding plot of dependence of the observed association rate (k_{obs}) on protein concentration, as recorded by BLI. B) Isotherms for the binding of the ssFUSE29 to FIR RRM1-RRM2 (NMR). S268 and S269 are two residues of RRM2 ~ 5 Å away from the proposed dimer interface which do not experience CSP in response to DNA binding. C) An X-ray structure of FIR

RRM1-RRM2 bound to the ssFUSE DNA (2QFJ). The FIR RRM1-RRM2 dimer is represented with a grey ribbon and residues whose H^N protons are less than 10 Å from the dimer interface are displayed in blue. The nucleotides visible in the X-ray structure are shown in orange. D) CSP of FIR RRM1-RRM2 resonances upon addition of ssFUSE29 or TGTGT pentamer. CSP by the ssFUSE29 (blue) largely overlap with CSP by the TGTGT pentamer (yellow) defining a common binding surface (green). Small differences are observed in the protein N-terminal helix, that is opposite the dimerisation area. These differences can be attributed to the transient contacts with additional nucleotides in the longer DNA. FIR RRM1-RRM2 does not dimerise on the ssFUSE29 DNA.

very similar, supporting the lack of dimerisation. The only (small) additional perturbations upon the binding of the longer DNA emerge in the N-terminal helix of RRM1. This helix is located on the protein side that is opposite the dimerisation interface, and the perturbations are most likely caused by transient and non-specific interactions of the DNA phosphates with the side-chains of two arginine residues. Furthermore, the rotational correlation time of the FIR RRM1-RRM2 – ssFUSE29 complex (τ_c = 18.8 ns) is too small for a ~ 50 kDa quasi-globular FIR RRM1-RRM2 dimer – ssFUSE29 complex (Figure 22). Interestingly, binding to the 29mer is only slightly stronger than binding to the 5mer. This difference can be easily accounted for by the presence of adjacent binding sites and is not consistent with a significant cooperativity expected for the binding of FIR dimer to the DNA.

In summary, the lack of CSP in RRM2 upon DNA binding, the analogous pattern of CSP in RRM1 upon titration with the 29mer and 5mer DNAs, the low rotational correlation time of the FIR RRM1-RRM2 – ssFUSE29 complex and the absence of a substantial cooperativity in the FIR binding to the longer DNA all argue against a significant DNA-mediated FIR RRM1-RRM2 dimerisation.

5.4 Simultaneous and independent recognition of ssFUSE and FBP Nbox peptide by FIR RRM1-RRM2

FIR RRM1-RRM2 binding interfaces for ssFUSE and FBP Nbox are located on the different domains and on the opposite sides of the protein (Figure 36A). In order to check if these two interactions are independent or if an allosteric mechanism exists the spectra of di-molecular complexes (protein – peptide and protein – DNA) were compared with a spectrum recorded upon addition of ssFUSE29 to the complex of FIR RRM1-RRM2 and FBP Nbox (Figure 36B). The CSPs observed in the tri-molecular titration combine the CSPs observed in separate titrations with ssDNA and the peptide indicating that FIR RRM1-RRM2 forms a tri-molecular complex with its binding partners. The possibility of allosteric cooperativity of peptide and DNA binding was excluded by comparing the affinity of specific interactions in tri- and di-molecular complexes. The affinity of FIR RRM1-RRM2 for the FBP Nbox peptide is not affected

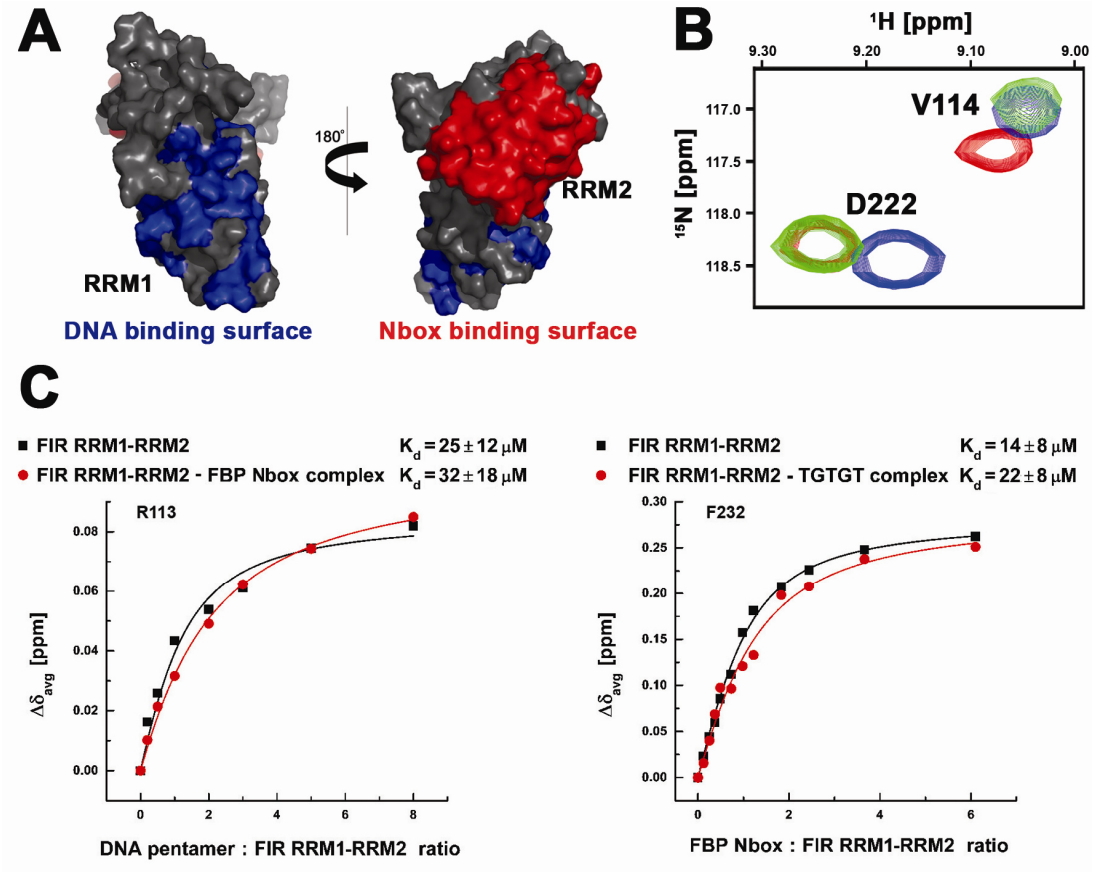

Figure 36 FIR RRM1-RRM2 simultaneously and independently interacts with ssDNA and FBP Nbox. A) Surface representation of FIR RRM1-RRM2. Residues showing significant CSP upon addition of ssFUSE29 and the FBP Nbox peptide are coloured in blue and red respectively. B) Superimposed ^{1}H-^{15}N correlation spectra show that FIR RRM1-RRM2 interacts independently with a ssFUSE29 and the FBP Nbox peptide. The spectra of FIR RRM1-RRM2+ssFUSE29, FIR RRM1-RRM2+FBP Nbox and FIR RRM1-RRM2+ssFUSE29+FBP Nbox are in blue, red and green respectively. A representative region containing one RRM1 resonance (V114) perturbed by DNA binding and one RRM2 resonance (D222) perturbed by peptide binding is displayed. The chemical shift changes of FIR RRM1-RRM2 peaks in the protein – DNA and protein – peptide complexes are additive in the tri-molecular complex, indicating independent binding. C) FBP and DNA bind independently to FIR RRM1-RRM2. Left – Superimposition of the isotherms for binding of TGTGT ssDNA to FIR RRM1-RRM2 (black) and to FIR RRM1-RRM2 – FBP Nbox complex (red). DNA binding to FIR RRM1-RRM2 is not affected by the presence of bound FBP Nbox peptide. Right – Superimposition of the isotherms for binding of the FBP Nbox peptide to FIR RRM1-RRM2 (black) and to FIR RRM1-RRM2 – TGTGT ssDNA (red). FBP Nbox peptide binding to FIR RRM1-RRM2 is not affected by the presence of bound DNA.

by the presence of bound TGTGT pentamer (K_d = 22 ± 8 μM in the presence of TGTGT pentamer *vs.* K_d = 14 ± 8 μM in the absence of DNA) nor does the presence of the bound peptide affect the affinity of FIR RRM1-RRM2 for the TGTGT pentamer (K_d = 32 ± 18 μM in the presence of the peptide *vs.* K_d = 25 ± 12 μM in the absence of the peptide) (Figure 36C).

5.5 FBP binding to ssFUSE

The analysis of BLI experiments described in this section has been performed with the help of Dr. Stephen R. Martin at the National Institute for Medical Research, London, UK. The two proteins (FBP Nbox-KH1-KH4 and FBP KH1-KH4) used in the experiments have been produced by David Hollingworth at the National Institute for Medical Research, London, UK.

The published data indicate that FBP protein binds to T-rich sequences (Benjamin et al., 2008; Braddock et al., 2002) as does FIR. The minimal size of FUSE-derived ssDNA necessary for FBP binding was determined with BLI. A number of biotinylated DNAs were immobilised on streptavidin-coated sensors, that were subsequently exposed to a range of FBP concentrations (0.5 to 8 nM) to study the association kinetics. The dissociation kinetics was followed by moving sensors with bound protein to the wells containing buffer. Kinetic analysis at high FBP concentration required double exponential fits to account for a minor component associated with non-specific binding. Initially, the experiments were performed with two different FUSE-derived ssDNAs (87mer – ssFUSE87 and 40mer – ssFUSE40) and two different FBP protein constructs (FBP Nbox-KH1-KH4 and FBP KH1-KH4) (Figure 37). The observed rate constants for the major part of the association phases (k_{obs}) were linearly dependent on FBP concentration. The analysis showed that the dissociation constants are identical for the binding to ssFUSE87 and ssFUSE40 (K_d ~ 1.7 nM) and revealed only a small (two-fold) difference in association and dissociation rate constants for the binding to ssFUSE87 and ssFUSE40 (k_{on} ~ 5.0 x 10^5 $M^{-1}s^{-1}$, k_{off} ~ 8.0 x 10^{-4} s^{-1} *vs.* k_{on} ~ 2.6 x 10^5 $M^{-1}s^{-1}$, k_{off} = 4.3 x 10^{-4} s^{-1}) (Table 9). This observed difference in the rate constants may reflect the accessibility and/or flexibility of the immobilised

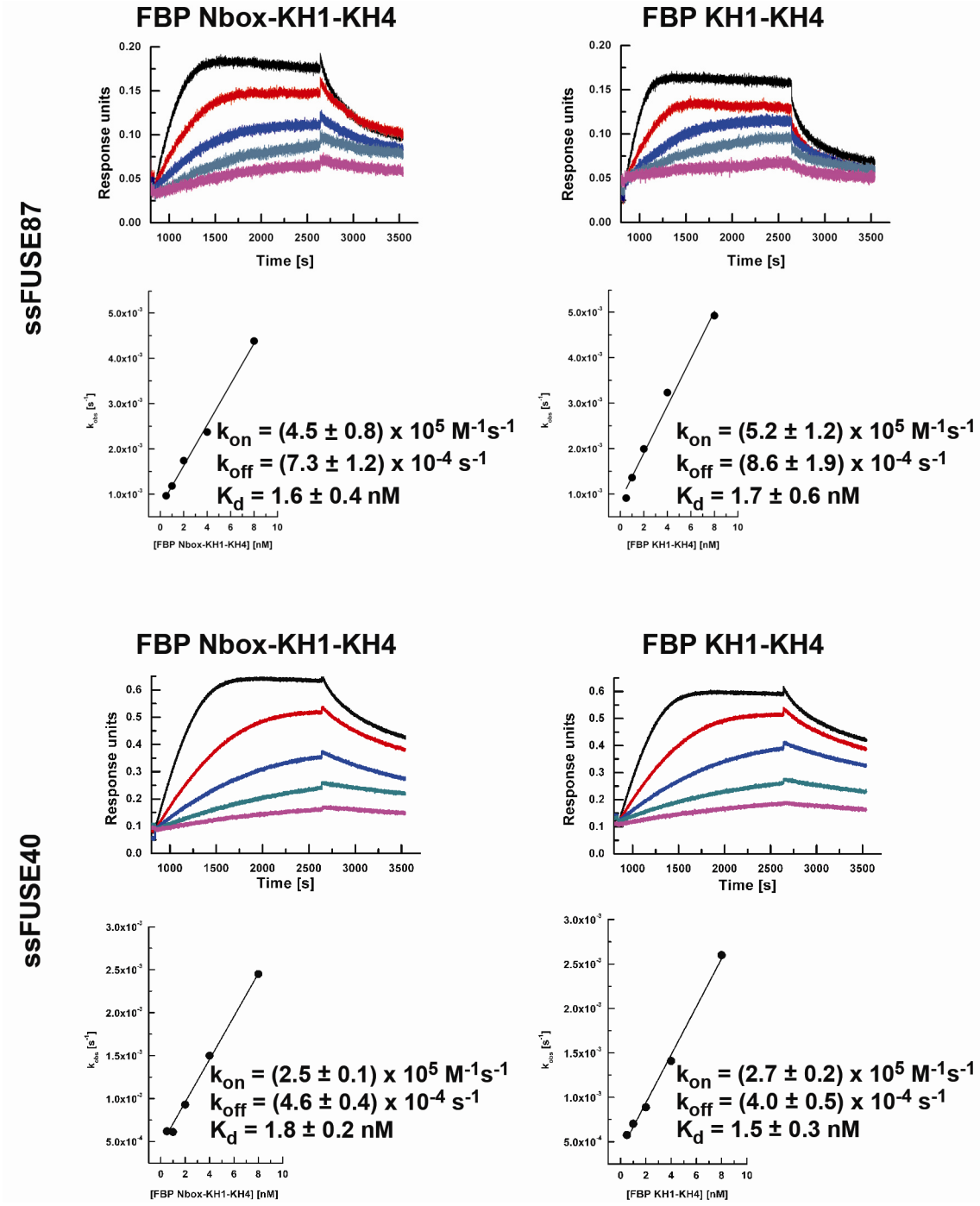

Figure 37 FBP binding to ssFUSE studied with BLI. Binding of increasing concentrations of FBP Nbox-KH1-KH4 (left) and FBP KH1-KH4 (right) to ssFUSE87 (top) and ssFUSE40 (bottom). Each panel shows kinetics curves and the dependence of the observed association rate (k_{obs}) on protein concentration together with obtained k_{on}, k_{off} and K_d values. The protein concentrations were 0.5 nM, 1 nM, 2 nM, 4 nM and 8 nM. The FBP Nbox does not contribute to the DNA binding and ssFUSE87 does not contain additional sequences compared to ssFUSE40 that are recognised by FBP.

Table 9 Kinetics parameters of two different FBP constructs binding to ssFUSE87 and ssFUSE40.

Oligonucleotide	Protein	k_{on} [$M^{-1}s^{-1}$]	k_{off} [s^{-1}]	K_d [nM]
ssFUSE87	FBP Nbox-KH1-KH4	$(4.5 \pm 0.8) \times 10^{5}$	$(7.3 \pm 1.2) \times 10^{-4}$	1.6 ± 0.4
	FBP KH1-KH4	$(5.2 \pm 1.2) \times 10^{5}$	$(8.6 \pm 1.9) \times 10^{-4}$	1.7 ± 0.6
ssFUSE40	FBP Nbox-KH1-KH4	$(2.5 \pm 0.1) \times 10^{5}$	$(4.6 \pm 0.4) \times 10^{-4}$	1.8 ± 0.2
	FBP KH1-KH4	$(2.7 \pm 0.2) \times 10^{5}$	$(4.0 \pm 0.5) \times 10^{-4}$	1.5 ± 0.3

DNA on the surface. Furthermore, the presence of Nbox in the FBP construct did not affect the binding to either of the two DNA (Table 9). We concluded that FBP Nbox does not participate in the DNA binding and that ssFUSE87 does not contain any additional FBP target sequences compared to ssFUSE40. Importantly, the defined DNA sequence is consistent with the FBP-binding sequence defined by Benjamin *et al.* (Benjamin et al., 2008).

Additional experiments with different length oligonucleotides derived from ssFUSE40 (Figure 38) and FBP KH1-KH4 protein were carried out to further confine the DNA sequence recognised by FBP. The truncations of ssFUSE40 from the 5' end (ssFUSE6-40, ssFUSE11-40 and ssFUSE16-40) did not significantly affect FBP binding (K_d = 1.9 ± 0.4 nM, K_d = 2.1 ± 0.6 nM and K_d = 2.6 ± 0.8 nM respectively), while 10-nucleotide (ssFUSE1-30) and 15-nucleotide (ssFUSE1-25) truncations from the 3' end diminished the binding ~ 15-fold (K_d = 27 ± 6 nM) and ~ 100-fold (K_d = ~ 100 nM) respectively (Figure 38 and Table 10). Thus, a ssFUSE that is 25-nucleotide long and that contains T-rich sequences is sufficient to bind FBP protein with high affinity.

5.6 FBP-mediated tethering of the FIR to the ssFUSE

The analysis of BLI experiments described in this section has been performed with the help of Dr. Stephen R. Martin at the National Institute for Medical Research, London, UK. The two proteins (FBP Nbox-KH1-KH4 and FBP KH1-KH4) used in the experiments have been produced by David Hollingworth at the National Institute for Medical Research, London, UK.

The effect of FBP Nbox on FIR binding to ssFUSE was studied using BLI. Sensors with immobilised ssFUSE40 were exposed first to FBP Nbox-KH1-KH4 and FBP KH1-KH4 at 100 nM, then to FBP solutions at the same concentration containing 150 nM FIR RRM1-RRM2, and finally to FBP solutions, again at the same concentration but without FIR (Figure 39). Significant binding of FIR was detected with FBP Nbox-KH1-KH4 (red curve) but no binding was detected with FBP KH1-KH4 (blue curve). The 10-fold increase in the FBP KH1-KH4 (up to 1 μM) did not lead to

```
ssFUSE40      5′-ATAATGTATATTCCCTCGGGATTTTTTATTTTGTGTTATT-3′

ssFUSE6-40          5′-GTATATTCCCTCGGGATTTTTTATTTTGTGTTATT-3′

ssFUSE11-40               5′-TTCCCTCGGGATTTTTTATTTTGTGTTATT-3′

ssFUSE16-40                     5′-TCGGGATTTTTTATTTTGTGTTATT-3′

ssFUSE1-35    5′-ATAATGTATATTCCCTCGGGATTTTTTATTTTGTG-3′

ssFUSE1-30    5′-ATAATGTATATTCCCTCGGGATTTTTTATT-3′

ssFUSE1-25    5′-ATAATGTATATTCCCTCGGGATTTT-3′
```

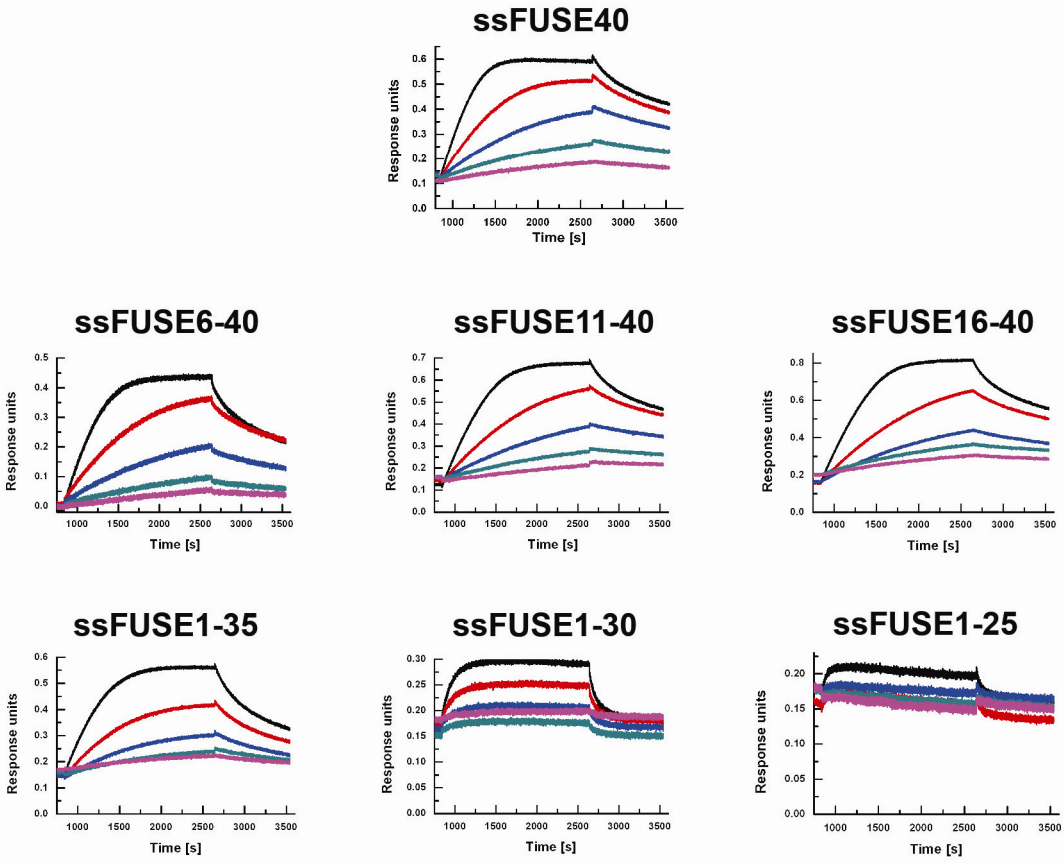

Figure 38 FBP binds to T-rich sequences. Top – a sequence alignment of oligonucleotides used to study FBP binding to ssDNA with BLI. Bottom – binding of increasing concentrations of FBP KH1-KH4 (0.5 nM, 1 nM, 2 nM, 4 nM and 8 nM) to different ssFUSE oligonucleotides. Deletions of the T-rich sequence at the 3’ end of ssFUSE40 significantly decrease FBP binding.

Table 10 Kinetics parameters of FBP KH1-KH4 binding to different ssFUSE-derived oligonucleotides.

Oligonucleotide	k_{on} [$M^{-1}s^{-1}$]	k_{off} [s^{-1}]	K_d [nM]
ssFUSE87	$(5.2 \pm 1.2) \times 10^{5}$	$(8.6 \pm 1.9) \times 10^{-4}$	1.7 ± 0.6
ssFUSE40	$(2.7 \pm 0.2) \times 10^{5}$	$(4.0 \pm 0.5) \times 10^{-4}$	1.5 ± 0.3
ssFUSE6-40	$(2.1 \pm 0.2) \times 10^{5}$	$(4.0 \pm 0.5) \times 10^{-4}$	1.9 ± 0.4
ssFUSE11-40	$(1.6 \pm 0.3) \times 10^{5}$	$(3.3 \pm 0.7) \times 10^{-4}$	2.1 ± 0.6
ssFUSE16-40	$(1.3 \pm 0.3) \times 10^{5}$	$(3.4 \pm 0.5) \times 10^{-4}$	2.6 ± 0.8
ssFUSE1-35	$(2.1 \pm 0.3) \times 10^{5}$	$(6.7 \pm 0.8) \times 10^{-4}$	3.2 ± 0.6
ssFUSE1-30	$(2.6 \pm 0.5) \times 10^{5}$	$(7.1 \pm 0.1) \times 10^{-3}$	27 ± 6
ssFUSE1-25	2.2×10^{5} [a]	$(2.2 \pm 0.3) \times 10^{-2}$	~ 100

[a] k_{on} assumed to be similar to all other oligonucleotides

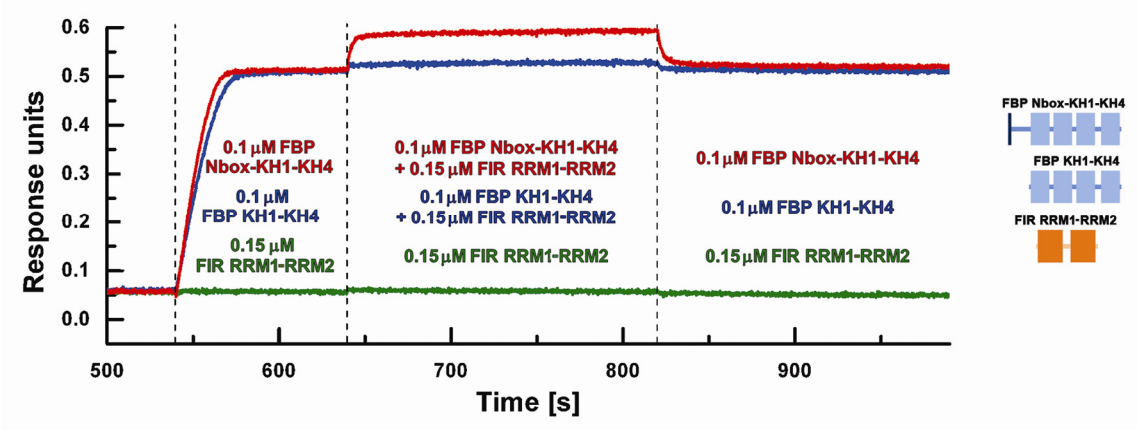

Figure 39 FBP Nbox recruits FIR to ssFUSE DNA. BLI binding assays show that the presence of the Nbox in FBP construct increases the affinity of FIR for the DNA. The BLI biosensors were derivatised with ssFUSE40 and exposed to different combinations of protein constructs (indicated on right), as reported in the figure. Response unit changes in three parallel experiments are displayed.

FIR binding (data not shown) and as expected, the binding of 150 nM FIR to ssFUSE40 could not be detected at this concentration (green curve). Unfortunately, a detailed kinetic analysis of FIR RRM1-RRM2 binding to the bound FBP Nbox-KH1-KH4 was impossible because of interference at high FIR concentrations from FIR binding to the DNA. Nevertheless, it is clear from inspection of the kinetic curves that the effect of FBP Nbox-KH1-KH4 is to decrease the dissociation rate constant for FIR from 0.68 s^{-1} to ~ 0.3 s^{-1} and to increase the association rate constant by a factor of 5-10. This would correspond to an increase in the affinity from K_d = 7 μM to K_d = ~ 0.3 – 0.7 μM.

Chapter 6

Discussion

6.1 Summary of the key findings

The recruitment of the FIR repressor to the ssFUSE – FBP complex is of central importance in the cell cycle dependent regulation of *c-myc*. The first two RRM domains of FIR function as a structural scaffold for the interactions between the three components of the FUSE–FBP–FIR system. In this thesis, I describe the structural and biophysical characterisation of the interactions between FIR and both FBP and ssFUSE, as well as exploring the FBP–ssFUSE interaction. This work has allowed me to propose a molecular model that captures important aspects of the regulation.

6.1.1 Structure of the first two RRM domains of FIR

The individual domains of FIR RRM1-RRM2 have a classical RRM fold, where two α-helices pack onto a four-stranded anti-parallel β-sheet. The two domains interact with each other, burying 1100 $Å^2$ of solvent accessible surface. The inter-domain interface encompasses the α-helices of RRM1 and the β-sheet of RRM2, and occludes the canonical RNA/DNA binding surface of RRM2. Two additional α-helical elements (in the N-terminus of the construct and in the linker between the two domains) further stabilise the relative orientation between RRM1 and RRM2.

The CD-monitored thermal denaturation of FIR RRM1-RRM2 shows a single transition point. This, together with the increase in stability of the domains when in the RRM1-RRM2 construct, emphasises that the two domains are joined in a single structural unit. Backbone relaxation data from NMR experiments confirm that the rotational correlation time (10.6 ns) is that expected for a compact ~ 20 kDa macromolecule (Maciejewski et al., 2000), consistent with FIR RRM1-RRM2 forming a single rigid unit. Furthermore, the T2 and hetNOE values show that significant intra-molecular motions are limited to the flexible loops and to the very ends of the FIR RRM1-RRM2 construct.

An inter-domain interaction between RRM domains has been reported for three other proteins in the absence of nucleic acid: hnRNPA1 (Shamoo et al., 1997; Xu et al., 1997), PTB (Vitali et al., 2006) and PRP24 (Bae et al., 2007) (Figure 40). In hnRNPA1,

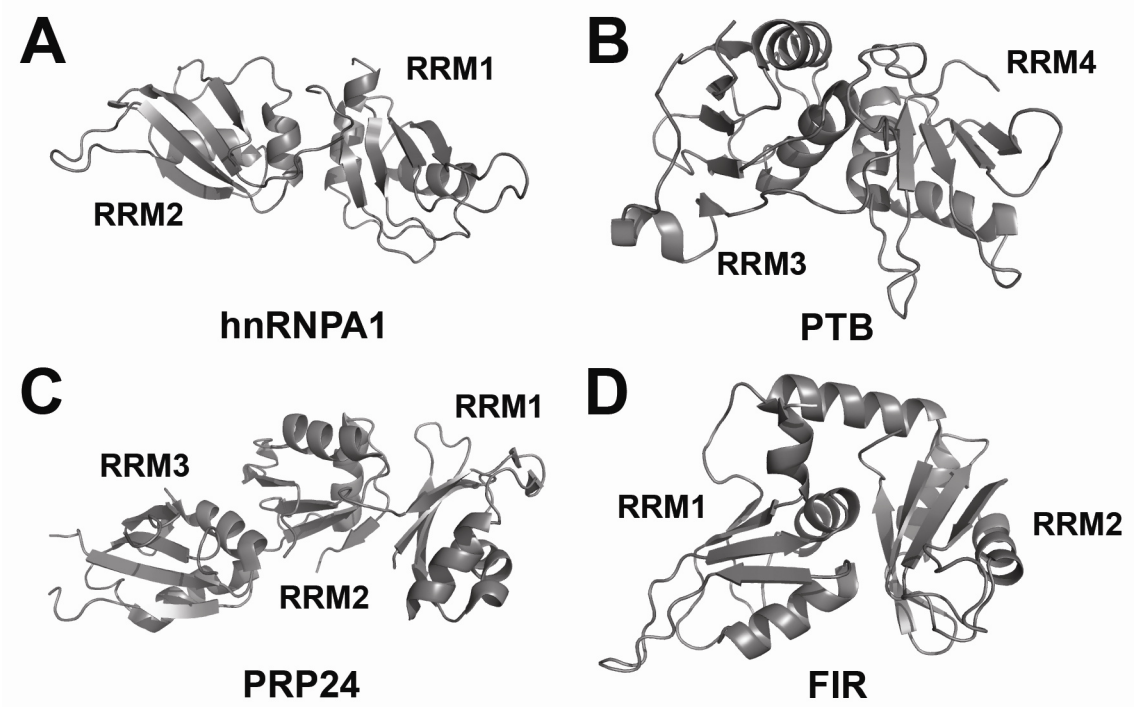

Figure 40 Comparison of known RRM–RRM interactions in the absence of nucleic acid. Ribbon representation of A) the first two RRM domains of hnRNPA1 (1HA1), B) RRM3 and RRM4 of PTB (2EVZ), C) the first three RRM domains of PRP24 (2GHP) and D) the first two RRM domains of FIR (this study).

the two RRM domains creates an extended β-sheet surface (composed of eight strands), that increases the affinity for RNA (Shamoo et al., 1997; Xu et al., 1997). In PTB, the two C-terminal RRM domains (RRM3 and RRM4) interact and this places the two nucleic acid binding sites far away from each other, implying either a bend in the RNA or binding of two separated sequences (Vitali et al., 2006). In PRP24, the first three RRMs are involved in inter-domain interactions (Bae et al., 2007). RRM2 of PRP24 interacts via its β-sheet surface with the β-sheet surface of RRM1 and with the α-helical surface of RRM3 (Bae et al., 2007). There are, however, no contacts between the RRM1 and RRM3 domains (Bae et al., 2007). Furthermore, NMR experiments show that all three domains of PRP24 contribute to the RNA binding, and therefore that the structural organisation of the domains probably changes upon binding to RNA (Bae et al., 2007). In FIR, as in PRP24, the inter-domain interaction between RRM1 and RRM2 occludes the nucleic acid binding surface of RRM2 (β-sheet surface). However, our NMR titration experiments and the X-ray structure of the FIR RRM1-RRM2 – ssFUSE complex (Crichlow et al., 2008) prove that RRM2 is not involved in the RNA/DNA binding. Therefore, a role of the pre-blocked organisation of FIR may be to appropriately position the protein within the FUSE regulatory complex. Analysis of our backbone relaxation data and comparison of FIR RRM1-RRM2 structures alone and Nbox-bound with the published ssDNA-bound structure (Crichlow et al., 2008) demonstrate that the protein fold and inter-domain arrangement are not modified by the interactions and further support a possible architectural function for the RRM1-RRM2 inter-domain interaction.

6.1.2 Structure of the FIR RRM1-RRM2 – FBP Nbox complex

The characterisation of the FBP Nbox by far-UV CD and NMR revealed that the peptide assumes an α-helical conformation in solution. The analysis of ^{15}N, $^{1}H^{\alpha}$, $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts defined the α-helix boundaries (D29 – I41), which are consistent with the ~ 40% α-helix content estimated by far-UV CD. Interestingly, it has been suggested that the Nbox forms an amphipathic helix, which binds to FIR via its hydrophobic face (Chung et al., 2006; Duncan et al., 1994). CSP analysis of the peptide

failed to confirm this hypothesis, because the perturbations were observed on both hydrophobic and hydrophilic faces of the putative helix during the titration with FIR RRM1-RRM2. However, this large effect can be attributed to the stabilisation of the Nbox in the helical conformation, that is confirmed by the observed chemical shift changes of the backbone (particularly for $C^{\alpha}H^{\alpha}$) resonances. Furthermore, φ and ψ dihedral angles derived from chemical shifts indicate that the helix extends to residues V27-I45 once the Nbox is in the complex. These 19 residues correspond to five turns of the helix with a length of ~ 27 Å, which is consistent with the interaction surface mapped on the FIR RRM2 (~ 15 x 25 Å). The interaction between the Nbox and RRM1-RRM2 is relatively weak (K_d ~ 15 μM) and the interface of the complex is dynamic, as reflected in the shorter-than-average T2 values of the resonances at the interface.

In the structure, the FBP Nbox peptide binds to FIR RRM1-RRM2 as an amphipathic α-helix with its hydrophobic face making contacts with a shallow hydrophobic groove on the α-helical side of the FIR RRM2 domain – consistent with CSP data on FIR RRM1-RRM2. Interestingly, there is no extensive inter-digitation of amino acid side-chains: the centre of the interface is occupied by alanines on the peptide side, while the longer hydrophobic side-chains of the peptide are loosely packed against the RRM2 surface. The structure of the complex explains the results of the published mutational analysis of the FBP–FIR interaction (Chung et al., 2006). In this work, a strong impairment of binding was observed for A34V and A38V mutants of the peptide that place a large hydrophobic chain in the centre of the inter-molecular interface and cause steric hindrance. Mutations of large hydrophobic residues (F31A or L35A), that are partially packed against the FIR RRM2, also attenuate the binding, but the effect is weaker than for the mutations of alanines, and can be attributed to the loss of specific hydrophobic interactions. In contrast, mutation of the loosely packed A42 (A42V) does not lead to a significant decrease in binding affinity. Recapitulating, the peptide–protein interface is designed to provide a low affinity (K_d ~ 15 μM), but highly specific recognition – a hallmark of the transient FIR–FBP interaction that is crucial for recruitment of FIR to the FUSE, but also has to be disrupted without impairing FUSE–FIR binding in order to allow FBP ejection from the DNA.

The change in the conformation of the Nbox upon complexation suggests that its propensity to form an α-helix could influence the interaction. Indeed, the peptide–

protein interaction is disrupted upon substitution of two solvent-exposed residues (K32 and D33) with prolines (a known helix-breaker), but not with alanines (Chung et al., 2006). Also mutation of A38 to another helix-breaking residue – glycine – impairs the binding (Chung et al., 2006). Finally, we found that FBP3 Nbox peptide, the weakest binder of the FBPs Nbox peptides, has also the lowest content of α-helix in solution. Therefore the sequence of the peptide defines the binding affinity both by creating specific inter-molecular side-chain – side-chain contacts and by determining the propensity of Nbox to assume the correct secondary structure.

The binding of FBP Nbox peptide in a shallow groove on the α-helical face of FIR RRM2 represents a novel mode of protein–protein interaction in the RRM family (Figure 41). In principle, both the α and β sides of RRM domains can be involved in the inter-protein interactions. In the Y14 – Mago – PYM tri-molecular complex, the RRM domain of Y14 interacts via its β-sheet surface with Mago, and via its β_2/β_3 loop with PYM (Bono et al., 2004). Similarly, UPF3b and p14 use the β-sheet face of RRM to bind respectively UPF2 and SF3b155 (Kadlec et al., 2004; Schellenberg et al., 2006). The Raver1 – vinculin interaction is directed by the binding of vinculin to the β_1/α_1, β_2/β_3 and α_2/β_4 loops in RRM1 of Raver1 (Lee et al., 2009). In the other known protein–protein complexes, RRM domains recognise their binding partner via the α-helical face. A short peptide of Raver1 (SLLGEPP) is placed across the α_1 and α_2 helices of PTB RRM2 (Rideau et al., 2006). In the U2B'' – U2A' complex, the α_1 helix of RRM in U2B'' packs against a β-sheet surface of the leucine-rich region in U2A' (Price et al., 1998). Finally, a separate distinct mode of recognition is represented by the UHM subfamily of RRM domains (Kielkopf et al., 2004). Here, the RXF motif (X being any residue) and acidic residues on the helical side of the UHM domain bind a conserved W and positive residues in the UHM ligand motif (ULM), as in U2AF35 – U2AF65 (Kielkopf et al., 2001), U2AF65 – SF1 (Selenko et al., 2003) and SPF45 – SF3b155 (Corsini et al., 2007) complexes. The recognition of FBP Nbox by the helical face of FIR RRM2 presents a novel mode of interaction because 1) the Nbox assumes an α-helical conformation and 2) the peptide binds in a quasi-parallel orientation to the α_1 helix of RRM2 rather than across both α_1 and α_2 helices.

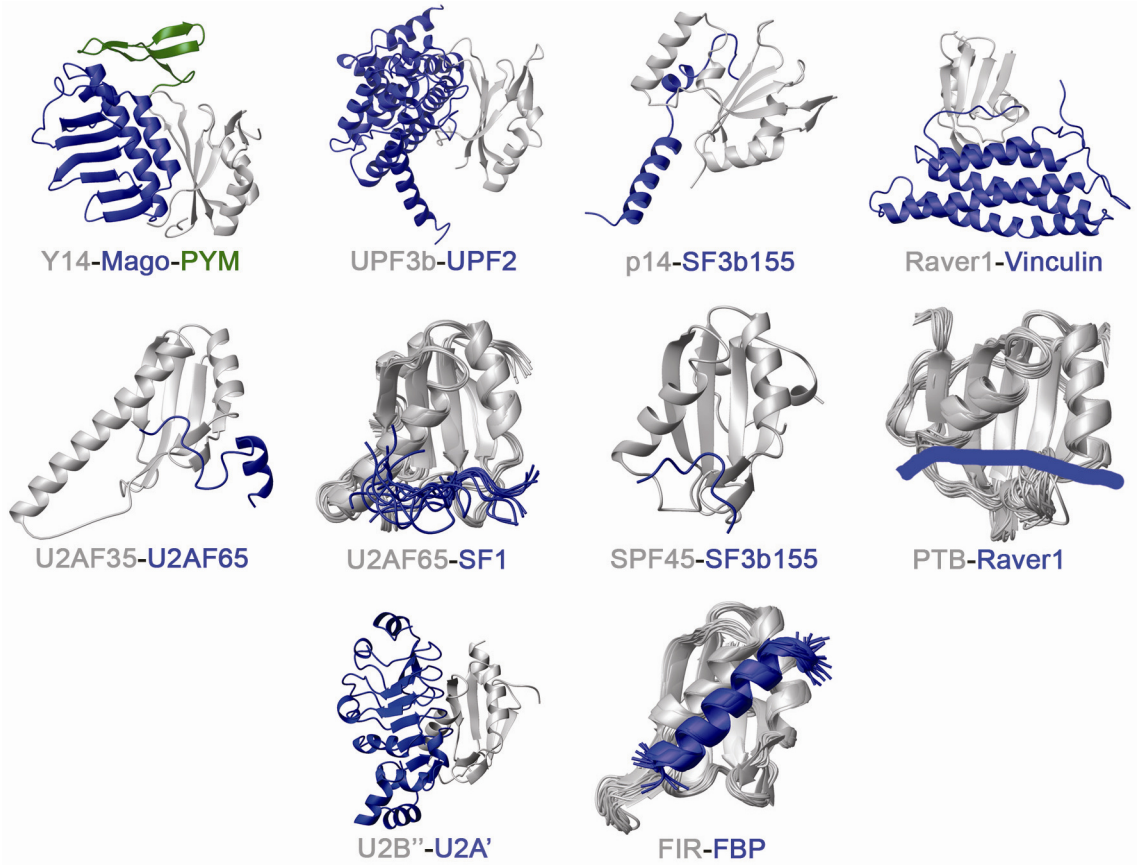

Figure 41 FIR–FBP interaction represents a novel recognition mode in the RRM family. Ribbon representations of the structures of protein–protein complexes between RRM domains (grey) and their binding partners (blue and green): Y14 – Mago – PYM (1RK8), UPF3b – UPF2 (1UW4), p14 – SF3b155 (2F9D), Raver1 – Vinculin (3H2U), U2AF35 – U2AF65 (1JMT), U2AF65 – SF1 (1O0P), SPF45 – SF3b155 (2PEH), U2B'' – U2A' (1A9N) and FIR – FBP (this study). In the PTB – Raver1 complex the Raver1 peptide position is reported (blue) on the structure of PTB RRM2 in the complex with RNA (2ADB), according to the published model (Rideau et al., 2006). U2AF35, U2AF65 and SPF45 RRM domains belong to the UHM subfamily, which binds a conserved W and positive residues in the ligand peptides.

6.1.3 Binding of FIR RRM1-RRM2 and FBP KH1-KH4 to ssFUSE

Both FIR and FBP interact with ssFUSE. The current model of FIR–ssFUSE recognition is based on the X-ray structure of FIR RRM1-RRM2 bound to a 25mer ssFUSE-derived sequence and SEC-LS/RI/UV (size exclusion chromatography coupled with light scattering, refractive index and ultraviolet absorbance) data on protein–DNA interaction (Crichlow et al., 2008). In the crystals, FIR dimerises via a large interface that involves both RRM domains (Crichlow et al., 2008). Both FIR molecules are bound to DNA, but poor electron density allowed full determination of only one bound nucleotide per protein molecule (Crichlow et al., 2008). Based on the asymmetry of DNA binding (CG for one FIR molecule and AT for the other) and solvent fraction analysis, it was concluded that the two FIR molecules, as a dimer, interact with one molecule of DNA (Crichlow et al., 2008). SEC-LS/RI/UV measurements revealed that FIR RRM1-RRM2 binds to ssDNA with a 2 : 1 stoichiometry and that it is monomeric in the absence of nucleic acid (Crichlow et al., 2008). However, these experiments did not show if the two binding proteins are actually interacting. Furthermore, the authors reported that analysis of the binding by fluorescence allowed them to measure two distinct K_ds (2 μM and 74 μM) for the two binding events (Crichlow et al., 2008). This is not consistent either with the SEC-LS/RI/UV data, where a fully bound complex is obtained at a concentration of ~ 60 μM of FIR RRM1-RRM2 or with a strong cooperativity that should be observed for the homodimerising molecules (Crichlow et al., 2008).

My results argue that no FIR RRM1-RRM2 dimerisation occurs upon binding of FIR RRM1-RRM2 to ssFUSE. We tested this by titrating FIR RRM1-RRM2 with a 29mer ssFUSE. CSP analysis maps the recognition surface to β-sheet and β_2/β_3 and α_2/β_4 loops of the RRM1 domain, including the residues interacting with the two modelled nucleotides in the X-ray structure of the FIR RRM1-RRM2 – ssFUSE complex (Crichlow et al., 2008). The stoichiometry of the binding is 2 : 1, consistent with the SEC-LS/RI/UV data (Crichlow et al., 2008) and reflecting the sequence preference of the FIR RRM1 – that we show here binds better to Ts and Gs. Indeed, two T-rich stretches are present in the tested ssFUSE29 (see Table 4). The quality of the NMR data and the relatively tight binding did not allow me to define the K_ds for the two binding events: nonetheless, a fast-to-intermediate exchange regime of

the resonances indicates that the dissociation constant is in the low micromolar range, consistent with K_d = 7 μM for the FIR RRM1-RRM2 – ssFUSE40 complex determined using BLI. Interestingly, CSP analysis with both the 29mer and a 5mer, whose sequence is included within the 29mer and has been indicated as high affinity by SIA, did not show any perturbations for residues at the putative dimer interface. Consistent with this, the rotational correlation time of the complex is shorter than what would be expected for a dimer–ssDNA complex. Finally, no substantial cooperativity exists for the binding of FIR to ssDNA – as the affinity for the 29mer is only marginally tighter than that for the 5mer. The observed difference in the affinity for different ssDNAs can be easily explained by the presence of quasi-equivalent binding sites within the longer DNA molecules. Taken together our data indicate that FIR does not dimerise on the DNA.

FBP has four nucleic acid-binding modules – KH1 to KH4 – and at least two of them are necessary for efficient binding of ssDNA (Duncan et al., 1994; Michelotti et al., 1996). Here, for the first time the interaction between FBP and ssFUSE has been studied quantitatively. The BLI experiments with two different constructs – FBP Nbox-KH1-KH4 and FBP KH1-KH4 – showed that the Nbox peptide does not contribute significantly to ssDNA binding and that the dissociation constants of the complexes are ~ 1.5 nM. Interestingly, 3' truncations of ssFUSE, that eliminate the T-rich sequence, result in decreased affinity for the FBP protein and show that this T-rich stretch is crucial for the high affinity protein–nucleic acid interaction. The three orders of magnitude difference between FIR and FBP affinities for ssDNA (7 μM *vs.* 1.5 nM) indicates that the faster recruitment of FBP to ssFUSE is driven by its higher affinity rather than by differences in protein concentration as proposed in a recent paper (Chung et al., 2006).

6.1.4 The role of the FBP Nbox peptide in FIR recruitment

It has been suggested, based on electrophoretic mobility shift assay (EMSA), that the FBP and FIR proteins form a tri-molecular complex on an 87mer ssFUSE (Benjamin et al., 2008). Here, we confirm that despite their similar sequence preference, the FBP and FIR proteins can bind simultaneously to a 40mer ssFUSE DNA. We also

show that FIR uses opposite surfaces in RRM1 and RRM2 to interact with both ssFUSE and FBP. Furthermore, the affinity of the FIR RRM1-RRM2 for ssDNA is increased 10- to 20-fold in the presence of the FBP construct containing Nbox and KH1 to KH4 domains (FBP Nbox-KH1-KH4), but not in the presence of the FBP construct without the Nbox (KH1-KH4). Therefore, we prove that the amphipathic peptide of the activator (FBP) acts as a recruitment element for the repressor (FIR).

The enhanced binding of FIR is not due to an allosteric effect linking FIR–ssFUSE and FIR–FBP interactions. In three-component NMR titrations the presence of bound ssDNA does not affect the affinity of FIR for the peptide and *vice versa*. The absence of an allosteric effect is important because ejection of FBP precedes ejection of FIR (Liu et al., 2006) and thus disruption of FIR–FBP interaction should not affect FIR interaction with ssFUSE DNA. Instead, the interaction between FIR and Nbox provides a physical tether that couples FIR–ssFUSE binding to the much stronger FBP–ssFUSE binding. Theoretically, if FIR–ssFUSE and FIR–FBP interactions were fully coupled the observed K_d of FIR binding should approach the product of the K_ds for the two complexes, i.e. $7 \times 10^{-6} \times 15 \times 10^{-6} = 1.05 \times 10^{-10}$ M. However, the Nbox is separated from the first KH domain of FBP by a 50 amino acid flexible linker, that acts as a decoupling element and reduces the effect by 3000-fold to 6500-fold (estimated K_d = 0.3 – 0.7 μM). This value is in agreement with the theoretically predicted 5500-fold decrease for systems with 50-residue unstructured linkers separating the two binding sites with K_ds of 7 μM and 15 μM (Shamoo et al., 1995).

6.2 A model of FUSE–FBP–FIR regulation

The existing model of FUSE-mediated regulation of *c-myc* (see Fig. 3) (Liu et al., 2006) explains the biological data, but does not provide the details of the mechanism at the molecular and structural level. My work explains how the architectural arrangement of the RRM domains of the FIR protein, together with the specific interactions between FUSE, FBP and FIR, contributes to the management of the *c-myc* transcription. My data clarify: 1) what is the driving force for the recruitment of

the activator (FBP) to the FUSE regulatory element; 2) how the repressor (FIR) is recruited to the same element and 3) how FUSE–FBP–FIR interactions may contribute to the switch between activation and repression of the *c-myc* transcription.

The transcription of *c-myc* is started in response to multiple inputs at the promoter, which increase the rate of the escape of the paused polymerase (Liu et al., 2006; Wierstra and Alves, 2008). The torsional stress generated leads to the melting of the FUSE element (Liu et al., 2006). Subsequently, FBP is recruited to boost the transcription. It was proposed that FBP binding precedes FIR binding, because FBP is five times more abundant in the cell than FIR (Liu et al., 2006). However, our data suggest that the primary determinant of the recruitment order is the difference in affinity of FBP and FIR for ssDNA with FBP binding three orders of magnitude more tightly than FIR to ssFUSE (K_d ~ 1.5 nM *vs.* K_d ~ 7 μM). The bound FBP establishes interactions with the basic transcriptional machinery at the *c-myc* promoter and further increases the rate of the polymerase escape (Liu et al., 2006).

Following the FBP-driven boost in the transcription, a repressor (FIR) is bound to the FBP–ssFUSE complex. This specific recruitment of FIR is central to the FUSE-mediated regulation because it defines the length of the FBP promoted unhindered activation of *c-myc*. We show that binding of FIR to the regulatory complex is dependent on the FIR interaction with the FBP Nbox peptide. We also demonstrate that this effect is purely due to physical tethering of FIR and not due to an allosteric effect. Lack of allostery is functionally important for the FUSE–FBP–FIR regulatory mechanism – it means that the FBP–FIR interaction can be disrupted without affecting FIR–ssFUSE affinity, and therefore, that the activator can leave the regulatory element before the repressor. Furthermore, the physical tethering of FIR is precisely tuned to allow enough *c-myc* expression to occur. A coupling that was too weak would slow down FIR binding and would cause detrimental overexpression of *c-myc*. Conversely, coupling that was too strong would lead to a rapid recruitment of FIR and immediate shut off of the expression (Figure 42). The actual coupling leads to 10- to 20-fold increase in the affinity of FIR for ssFUSE. This is achieved by an amphipathic helix of FBP that binds to the RRM2 surface of FIR with a moderate affinity (K_d ~ 15 μM) and by a ~ 50 amino acid flexible linker that links this helix to the first KH domain – that acts as a decoupling element reducing the coupling effect ~ 5000-fold.

FIR acts as a repressor and it has been shown that it remains associated with the ssFUSE, after the ejection of FBP protein (Liu et al., 2006). However, the molecular

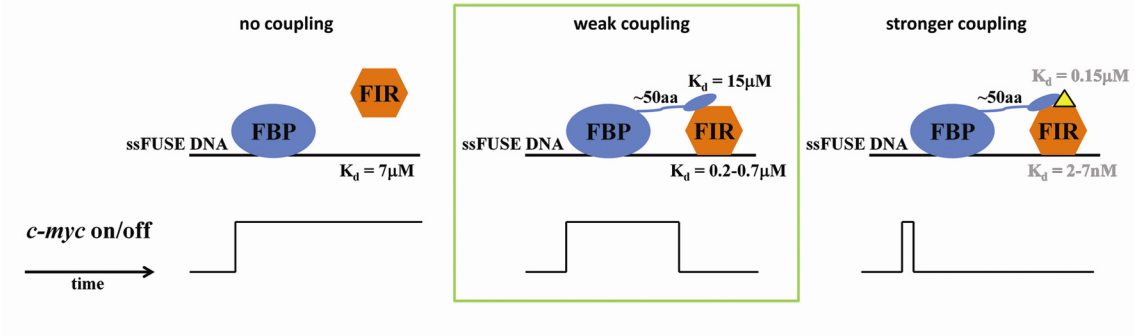

Figure 42 Three different modes of FBP–FIR coupling and their effects on FIR binding and *c-myc* transcription. Left – In the absence of any coupling between FBP and FIR the affinity of FIR for the ssFUSE (K_d ~ 7 µM) would be too low for FIR to bind the ssFUSE. FBP activation of *c-myc* transcription (bottom) would therefore continue unperturbed. Middle – Boxed, the physiological scenario. The weak binding of the Nbox to FIR RRM2 and the 50-amino acid linker between the Nbox and KH1 are responsible for the physiologically weak coupling between FBP and FIR. This coupling is necessary to create the required peak in *c-myc* expression (bottom) and regulate the cell cycle. Right – A stronger coupling between FBP and FIR created by, for example, a small molecular weight compound binding at the interface between the two molecules would increase ~ 100-fold the apparent affinity of FIR for the FBP–ssFUSE complex and would speed FIR recruitment and lead to a shorter expression of *c-myc* (bottom).

mechanism of the switch between activation and repression is poorly understood. Recently, a model was proposed where FIR dimerisation is linked to the ejection of FBP from the FUSE regulatory element (Crichlow et al., 2008). In this model, two molecules of FIR bind to the ssFUSE and dimerise. This leads to the looping of the DNA and structural rearrangements within the whole regulatory complex, which ultimately cause ejection of the FBP protein (Figure 43) (Crichlow et al., 2008). The model was based on the observed dimerisation of FIR on the ssFUSE in the X-ray structure. However, we show here that dimerisation does not take place in solution and therefore cannot be the driving force for the switch. Furthermore, the model is not consistent with our proposed mechanism of the recruitment, because FBP is able to bind and recruit only one molecule of FIR. Our data show that the FIR–FBP interaction does not allosterically affect the FIR–ssFUSE binding. Also, even though FIR binds to ssFUSE three orders of magnitude more weakly than FBP does, it is FIR that remains associated with the DNA for longer in the cell (Liu et al., 2006). Therefore, the FIR–ssFUSE interaction in the absence of FBP has to be stabilised by other components of the regulatory complex. A possible candidate is the XPB subunit of TFIIH transcription factor, that is involved in the initiation of *c-myc* transcription, and that interacts with the N-terminal part of FIR. Interestingly, FBP also interacts with XPB and it is tempting to speculate that FIR binding to ssFUSE and its subsequent interaction with XPB leads to a rearrangement of the TFIIH complex and thereby to the disruption of FBP–TFIIH interaction and ejection of FBP. Such a scenario seems more likely than a direct competition of FIR and FBP for the same binding site, as FBP and FIR interact with the XPB subunit of TFIIH using very different protein sequences (Liu et al., 2001; Liu et al., 2000). It is important to notice that FIR uses two RRM domains to interact with FBP and ssFUSE. This is unlikely to be imposed by the requirement for the lack of an allosteric effect as it was reported that a single RRM domain (RRM2 of PTB) can mediate protein–protein and protein–nucleic acid interaction without any allostery (Rideau et al., 2006). Therefore, the pre-formed single-unit organisation of two RRM domains within FIR protein may provide a special scaffold for the interactions, leading to the large rearrangements within the complex and causing FBP ejection.

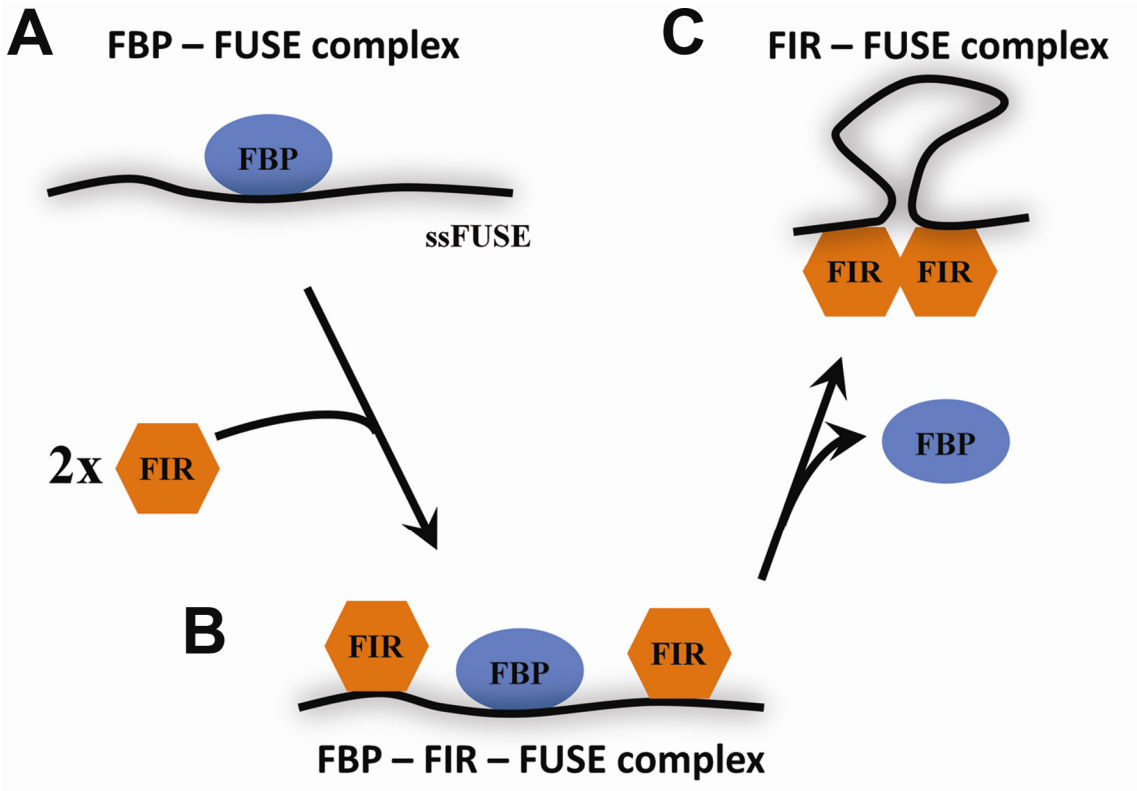

Figure 43 Model of the on/off switch in the transcriptional regulation of *c-myc* proposed by Crichlow *et al.* (adapted from Crichlow et al., 2008). A) FBP binds first to the FUSE and activates *c-myc* transcription. B) Subsequently two molecules of FIR bind to the FUSE and C) by dimerisation, looping out the DNA and conformational rearrangements of the complex eject FBP protein and lead to the repression of *c-myc* expression.

6.3 Perspectives

6.3.1 FUSE–FBP–FIR system as a drug target

In most cancer pathologies where *c-myc* function is disrupted the effect is caused by deregulation of the gene expression (and not by mutations in the c-MYC protein) (Oster et al., 2002; Vita and Henriksson, 2006). Because of the huge complexity of the transcriptional regulation of *c-myc* many of these pathologies very likely arise from the mutations, that directly or indirectly affect *c-myc* promoter regulation (Wierstra and Alves, 2008). The identification of such mutations and subsequent treatment would require a very personalised therapy approach, which would be both expensive and long. Importantly, FUSE–FBP–FIR acts independently from the single inputs converging to the *c-myc* promoter and therefore represents an attractive target for drugs. One possible intervention is the blockage of FBP binding to ssFUSE, for example with benzoylanthranilic acid inhibitors (Huth et al., 2004). However, this blockage may have some serious side-effects because the interactions of FBP with other DNA and/or RNA sequences will also be affected, thereby disrupting the multiple functions of FBP. Indeed, siRNA experiments confirmed that FBP proteins regulate expression of a broad range of genes (Chung et al., 2006).

My results suggest an alternative approach that would speed up the recruitment of FIR to the activated *c-myc* promoter and prolong FIR-mediated repression. The structure of the FIR RRM1-RRM2 – FBP Nbox complex and the studies on their interaction suggest that a low molecular weight ligand could be selected that would stabilise the interaction between the two partners (Figure 42). Therefore the recruitment of FIR should occur earlier and should significantly decrease FBP-mediated activation of the *c-myc* proto-oncogene. Interestingly, a similar enhancement of protein–protein interaction has been recently described for 14-3-3 protein and its target peptides (Ottmann et al., 2009). Here the complex is stabilised by cotylenin A – a fungal diterpene glycoside, that has pro-differentiative activity and therefore is a potential anti-cancer agent (Ottmann et al., 2009).

6.3.2 FUSE–FBP–FIR system as a research tool

A highly controlled gene expression is desired both as a research and therapeutic tool (Guo et al., 2008; Yamamoto et al., 2001). Many different approaches have been employed to drive expression of a target gene at the right level, right place and right time (Guo et al., 2008). Broad applications of these systems had been preceded by years of studies to define necessary elements, determine a mode of action and evaluate usefulness and safety (Guo et al., 2008; Yamamoto et al., 2001). Furthermore, in spite of the fact that all the systems are derived from natural counterparts they were often modified and joined in different combinations to broaden a spectrum of reachable profiles of expression (Guo et al., 2008; Yamamoto et al., 2001). For example, usage of heterologous promoters allows the expression of a desired gene ubiquitously (e.g. CMV, EF1α promoters) or limits it to a particular tissue, developmental stage or cell cycle stage (e.g. the albumin promoter limits expression to hepatocytes) (Guo et al., 2008). The promoters may be also responsive to environmental conditions such as heat shock, hypoxia, radiation, chemotherapy and viral infection (Guo et al., 2008). Particular attention is given to small-molecule-inducible expression systems that allow a regulated gene to be switched on/off at strictly controlled times, and thus, provide a possibility to tightly control level of expression, and above all, increase the safety as the gene can be switched off when necessary (Guo et al., 2008).

FUSE–FBP–FIR represents an attractive heterologous system that can be employed to create a characteristic peak of expression. The principal advantage of this regulation is the ability to tightly control the level and period of expression via self-inhibition. The mechanism can be used in combination with a highly regulated promoter (as in *c-myc* promoter) to create a chimeric regulatory unit that combines a strict control and an efficient up-regulation. However, virtually any promoter can be coupled with the FUSE–FBP–FIR system, because it responds to ongoing transcription (Liu et al., 2006). In fact, the mechanism has been already used to reprogram a response of the metallothionein IIA promoter, which initiates transcription upon addition of Zn^{2+} (Liu et al., 2006). Furthermore, the system can be tuned to obtain a desired shape of the peak, for example by varying a distance between the promoter and the FUSE sequence (Liu et al., 2006). Additional studies are necessary to fully evaluate the regulatory capabilities of the FUSE–FBP–FIR system.

References

Amati, B. (2004). Myc degradation: dancing with ubiquitin ligases. Proc Natl Acad Sci U S A *101*, 8843-8844.

Avigan, M.I., Strober, B., and Levens, D. (1990). A far upstream element stimulates c-myc expression in undifferentiated leukemia cells. J Biol Chem *265*, 18538-18545.

Bae, E., Reiter, N.J., Bingman, C.A., Kwan, S.S., Lee, D., Phillips, G.N., Jr., Butcher, S.E., and Brow, D.A. (2007). Structure and interactions of the first three RNA recognition motifs of splicing factor prp24. J Mol Biol *367*, 1447-1458.

Barreau, C., Paillard, L., and Osborne, H.B. (2005). AU-rich elements and associated factors: are there unifying principles? Nucleic Acids Res *33*, 7138-7150.

Bartels, C., Xia, T.-h., Billeter, M., Güntert, P., and Wüthrich, K. (1995). The program XEASY for computer-supported NMR spectral analysis of biological macromolecules. Journal of Biomolecular NMR *6*, 1-10.

Bax, A., and Grzesiek, S. (1993). Methodological advances in protein NMR. Acc Chem Res *26*, 131-138.

Bazar, L., Meighen, D., Harris, V., Duncan, R., Levens, D., and Avigan, M. (1995). Targeted melting and binding of a DNA regulatory element by a transactivator of c-myc. J Biol Chem *270*, 8241-8248.

Benjamin, L.R., Chung, H.J., Sanford, S., Kouzine, F., Liu, J., and Levens, D. (2008). Hierarchical mechanisms build the DNA-binding specificity of FUSE binding protein. Proc Natl Acad Sci U S A *105*, 18296-18301.

Bernstein, P.L., Herrick, D.J., Prokipcak, R.D., and Ross, J. (1992). Control of c-myc mRNA half-life in vitro by a protein capable of binding to a coding region stability determinant. Genes Dev *6*, 642-654.

Beuth, B., Garcia-Mayoral, M.F., Taylor, I.A., and Ramos, A. (2007). Scaffold-independent analysis of RNA-protein interactions: the Nova-1 KH3-RNA complex. J Am Chem Soc *129*, 10205-10210.

Bono, F., Ebert, J., Unterholzner, L., Guttler, T., Izaurralde, E., and Conti, E. (2004). Molecular insights into the interaction of PYM with the Mago-Y14 core of the exon junction complex. EMBO Rep *5*, 304-310.

Braddock, D.T., Louis, J.M., Baber, J.L., Levens, D., and Clore, G.M. (2002). Structure and dynamics of KH domains from FBP bound to single-stranded DNA. Nature *415*, 1051-1056.

Briata, P., Forcales, S.V., Ponassi, M., Corte, G., Chen, C.Y., Karin, M., Puri, P.L., and Gherzi, R. (2005). p38-dependent phosphorylation of the mRNA decay-promoting factor KSRP controls the stability of select myogenic transcripts. Mol Cell *20*, 891-903.

Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., *et al.* (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. Acta Crystallogr D Biol Crystallogr *54*, 905-921.

Card, P.B., and Gardner, K.H. (2005). Identification and optimization of protein domains for NMR studies. Methods Enzymol *394*, 3-16.

Cavanagh, J., Fairbrother, W. J., Palmer, A. G. III, Rance, M., Skelton, N. J. (2007). Protein NMR Spectroscopy (Second Edition) (London: Elsevier Inc.).

Chung, H.J., and Levens, D. (2005). c-myc expression: keep the noise down! Mol Cells *20*, 157-166.

Chung, H.J., Liu, J., Dundr, M., Nie, Z., Sanford, S., and Levens, D. (2006). FBPs are calibrated molecular tools to adjust gene expression. Mol Cell Biol *26*, 6584-6597.

Clery, A., Blatter, M., and Allain, F.H. (2008). RNA recognition motifs: boring? Not quite. Curr Opin Struct Biol *18*, 290-298.

Cole, M.D., and Cowling, V.H. (2008). Transcription-independent functions of MYC: regulation of translation and DNA replication. Nat Rev Mol Cell Biol *9*, 810-815.

Collins, S., and Groudine, M. (1982). Amplification of endogenous myc-related DNA sequences in a human myeloid leukaemia cell line. Nature *298*, 679-681.

Cornilescu, G., Delaglio, F., and Bax, A. (1999). Protein backbone angle restraints from searching a database for chemical shift and sequence homology. J Biomol NMR *13*, 289-302.

Corsini, L., Bonnal, S., Basquin, J., Hothorn, M., Scheffzek, K., Valcarcel, J., and Sattler, M. (2007). U2AF-homology motif interactions are required for alternative splicing regulation by SPF45. Nat Struct Mol Biol *14*, 620-629.

Corsini, L., Hothorn, M., Stier, G., Rybin, V., Scheffzek, K., Gibson, T.J., and Sattler, M. (2009). Dimerization and protein binding specificity of the U2AF homology motif of the splicing factor Puf60. J Biol Chem *284*, 630-639.

Crichlow, G.V., Zhou, H., Hsiao, H.H., Frederick, K.B., Debrosse, M., Yang, Y., Folta-Stogniew, E.J., Chung, H.J., Fan, C., De la Cruz, E.M., *et al.* (2008). Dimerization of FIR upon FUSE DNA binding suggests a mechanism of c-myc inhibition. EMBO J *27*, 277-289.

Dang, C.V., O'Donnell, K.A., Zeller, K.I., Nguyen, T., Osthus, R.C., and Li, F. (2006). The c-Myc target gene network. Semin Cancer Biol *16*, 253-264.

Davis-Smyth, T., Duncan, R.C., Zheng, T., Michelotti, G., and Levens, D. (1996). The far upstream element-binding proteins comprise an ancient family of single-strand DNA-binding transactivators. J Biol Chem *271*, 31679-31687.

Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J., and Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR *6*, 277-293.

Dempsey, C.E. (2001). Hydrogen exchange in peptides and proteins using NMR spectroscopy. Progress in Nuclear Magnetic Resonance Spectroscopy *39*, 135-170.

Duncan, R., Bazar, L., Michelotti, G., Tomonaga, T., Krutzsch, H., Avigan, M., and Levens, D. (1994). A sequence-specific, single-strand binding protein activates the far upstream element of c-myc and defines a new DNA-binding motif. Genes Dev *8*, 465-480.

Duncan, R., Collins, I., Tomonaga, T., Zhang, T., and Levens, D. (1996). A unique transactivation sequence motif is found in the carboxyl-terminal domain of the single-strand-binding protein FBP. Mol Cell Biol *16*, 2274-2282.

Edison, A.S., Weinhold, F., Westler, W.M., and Markley, J.L. (1994). Estimates of phi and psi torsion angles in proteins from one-, two- and three-bond nuclear spin-spin couplings: application to staphylococcal nuclease. J Biomol NMR *4*, 543-551.

Eilers, M., and Eisenman, R.N. (2008). Myc's broad reach. Genes Dev *22*, 2755-2766.

Feeney, J., Batchelor, J.G., Albrand, J.P., and Roberts, G.C.K. (1979). The effects of intermediate exchange processes on the estimation of equilibrium constants by NMR. Journal of Magnetic Resonance (1969) *33*, 519-529.

Fesik, S.W., and Zuiderweg, E.R.P. (1988). Heteronuclear three-dimensional nmr spectroscopy. A strategy for the simplification of homonuclear two-dimensional NMR spectra Journal of Magnetic Resonance *78*, 588-593.

Gardner, K.H., and Kay, L.E. (1998). The use of 2H, 13C, 15N multidimensional NMR to study the structure and dynamics of proteins. Annu Rev Biophys Biomol Struct *27*, 357-406.

Gherzi, R., Lee, K.Y., Briata, P., Wegmuller, D., Moroni, C., Karin, M., and Chen, C.Y. (2004). A KH domain RNA binding protein, KSRP, promotes ARE-directed mRNA turnover by recruiting the degradation machinery. Mol Cell *14*, 571-583.

Goddard, T.D., and Kneller, D.G. (2004). Sparky (computer program). (San Francisco, University of California).

Goga, A., Yang, D., Tward, A.D., Morgan, D.O., and Bishop, J.M. (2007). Inhibition of CDK1 as a potential therapy for tumors over-expressing MYC. Nat Med *13*, 820-827.

Grzesiek, S., Cordier, F., and Dingley, A.J. (2001). Scalar couplings across hydrogen bonds. Methods Enzymol *338*, 111-133.

Güntert, P. (2003). Automated NMR protein structure calculation. Progress in Nuclear Magnetic Resonance Spectroscopy *43*, 105-125.

Guo, Z.S., Li, Q., Bartlett, D.L., Yang, J.Y., and Fang, B. (2008). Gene transfer: the challenge of regulated gene expression. Trends Mol Med *14*, 410-418.

Habeck, M., Rieping, W., Linge, J.P., and Nilges, M. (2004). NOE assignment with ARIA 2.0: the nuts and bolts. Methods Mol Biol *278*, 379-402.

He, L., Liu, J., Collins, I., Sanford, S., O'Connell, B., Benham, C.J., and Levens, D. (2000). Loss of FBP function arrests cellular proliferation and extinguishes c-myc expression. EMBO J *19*, 1034-1044.

Hernandez, G., and LeMaster, D.M. (2009). NMR analysis of native-state protein conformational flexibility by hydrogen exchange. Methods Mol Biol *490*, 285-310.

Herold, S., Herkert, B., and Eilers, M. (2009). Facilitating replication under stress: an oncogenic function of MYC? Nat Rev Cancer *9*, 441-444.

Huth, J.R., Yu, L., Collins, I., Mack, J., Mendoza, R., Isaac, B., Braddock, D.T., Muchmore, S.W., Comess, K.M., Fesik, S.W., *et al.* (2004). NMR-driven discovery of benzoylanthranilic acid inhibitors of far upstream element binding protein binding to the human oncogene c-myc promoter. J Med Chem *47*, 4851-4857.

Ioannidis, P., Mahaira, L., Papadopoulou, A., Teixeira, M.R., Heim, S., Andersen, J.A., Evangelou, E., Dafni, U., Pandis, N., and Trangas, T. (2003). CRD-BP: a c-Myc mRNA stabilizing protein with an oncofetal pattern of expression. Anticancer Res *23*, 2179-2183.

Kadlec, J., Izaurralde, E., and Cusack, S. (2004). The structural basis for the interaction between nonsense-mediated mRNA decay factors UPF2 and UPF3. Nat Struct Mol Biol *11*, 330-337.

Kannt, A., Young, S., and Bendall, D.S. (1996). The role of acidic residues of plastocyanin in its interaction with cytochrome f. Biochimica Et Biophysica Acta-Bioenergetics *1277*, 115-126.

Kay, L.E., Torchia, D.A., and Bax, A. (1989). Backbone dynamics of proteins as studied by 15N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. Biochemistry *28*, 8972-8979.

Kay, L.E., Xu, G.Y., Singer, A.U., Muhandiram, D.R., and Formankay, J.D. (1993). A gradient-enhanced HCCH-TOCSY experiment for recording side-chain 1H and 13C correlations in H2O samples of proteins. Journal of Magnetic Resonance, Series B *101*, 333-337.

Kenneth, N.S., and White, R.J. (2009). Regulation by c-Myc of ncRNA expression. Curr Opin Genet Dev *19*, 38-43.

Kielkopf, C.L., Lucke, S., and Green, M.R. (2004). U2AF homology motifs: protein recognition in the RRM world. Genes Dev *18*, 1513-1526.

Kielkopf, C.L., Rodionova, N.A., Green, M.R., and Burley, S.K. (2001). A novel peptide recognition mode revealed by the X-ray structure of a core U2AF35/U2AF65 heterodimer. Cell *106*, 595-605.

Knoepfler, P.S. (2007). Myc goes global: new tricks for an old oncogene. Cancer Res *67*, 5061-5063.

Koradi, R., Billeter, M., and Wuthrich, K. (1996). MOLMOL: a program for display and analysis of macromolecular structures. J Mol Graph *14*, 51-55, 29-32.

Kouzine, F., and Levens, D. (2007). Supercoil-driven DNA structures regulate genetic transactions. Front Biosci *12*, 4409-4423.

Kouzine, F., Liu, J., Sanford, S., Chung, H.J., and Levens, D. (2004). The dynamic response of upstream DNA to transcription-generated torsional stress. Nat Struct Mol Biol *11*, 1092-1100.

Kouzine, F., Sanford, S., Elisha-Feil, Z., and Levens, D. (2008). The functional response of upstream DNA to dynamic supercoiling in vivo. Nat Struct Mol Biol *15*, 146-154.

Kroll, T.T., Zhao, W.M., Jiang, C., and Huber, P.W. (2002). A homolog of FBP2/KSRP binds to localized mRNAs in Xenopus oocytes. Development *129*, 5609-5619.

Laskowski, R.A., Rullmannn, J.A., MacArthur, M.W., Kaptein, R., and Thornton, J.M. (1996). AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J Biomol NMR *8*, 477-486.

Lee, J.H., Rangarajan, E.S., Yogesha, S.D., and Izard, T. (2009). Raver1 interactions with vinculin and RNA suggest a feed-forward pathway in directing mRNA to focal adhesions. Structure *17*, 833-842.

Levens, D. (2002). Disentangling the MYC web. Proc Natl Acad Sci U S A *99*, 5757-5759.

Linge, J.P., O'Donoghue, S.I., and Nilges, M. (2001). Automated assignment of ambiguous nuclear overhauser effects with ARIA. Methods Enzymol *339*, 71-90.

Linge, J.P., Williams, M.A., Spronk, C.A., Bonvin, A.M., and Nilges, M. (2003). Refinement of protein structures in explicit solvent. Proteins *50*, 496-506.

Liu, J., Akoulitchev, S., Weber, A., Ge, H., Chuikov, S., Libutti, D., Wang, X.W., Conaway, J.W., Harris, C.C., Conaway, R.C., *et al.* (2001). Defective interplay of activators and repressors with TFIH in xeroderma pigmentosum. Cell *104*, 353-363.

Liu, J., He, L., Collins, I., Ge, H., Libutti, D., Li, J., Egly, J.M., and Levens, D. (2000). The FBP interacting repressor targets TFIIH to inhibit activated transcription. Mol Cell *5*, 331-341.

Liu, J., Kouzine, F., Nie, Z., Chung, H.J., Elisha-Feil, Z., Weber, A., Zhao, K., and Levens, D. (2006). The FUSE/FBP/FIR/TFIIH system is a molecular machine programming a pulse of c-myc expression. EMBO J *25*, 2119-2130.

Maciejewski, M.W., Liu, D., Prasad, R., Wilson, S.H., and Mullen, G.P. (2000). Backbone dynamics and refined solution structure of the N-terminal domain of DNA polymerase beta. Correlation with DNA binding and dRP lyase activity. J Mol Biol *296*, 229-253.

Marion, D., Driscoll, P.C., Kay, L.E., Wingfield, P.T., Bax, A., Gronenborn, A.M., and Clore, G.M. (1989). Overcoming the overlap problem in the assignment of 1H NMR spectra of larger proteins by use of three-dimensional heteronuclear 1H-15N Hartmann-Hahn-multiple quantum coherence and nuclear Overhauser-multiple quantum coherence spectroscopy: application to interleukin 1 beta. Biochemistry *28*, 6150-6156.

Maris, C., Dominguez, C., and Allain, F.H. (2005). The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. FEBS J *272*, 2118-2131.

Markley, J.L., Ulrich, E.L., Westler, W.M., and Volkman, B.F. (2003). Macromolecular structure determination by NMR spectroscopy. Methods Biochem Anal *44*, 89-113.

Martin, S.R., and Schilstra, M.J. (2008). Circular dichroism and its application to the study of biomolecules. Methods Cell Biol *84*, 263-293.

Matsushita, K., Tomonaga, T., Kajiwara, T., Shimada, H., Itoga, S., Hiwasa, T., Kubo, S., Ochiai, T., Matsubara, H., and Nomura, F. (2009). c-myc suppressor FBP-interacting repressor for cancer diagnosis and therapy. Front Biosci *14*, 3401-3408.

Matsushita, K., Tomonaga, T., Shimada, H., Shioya, A., Higashi, M., Matsubara, H., Harigaya, K., Nomura, F., Libutti, D., Levens, D., and Ochiai, T. (2006). An essential role of alternative splicing of c-myc suppressor FUSE-binding protein-interacting repressor in carcinogenesis. Cancer Res *66*, 1409-1417.

Metzler, W.J., Constantine, K.L., Friedrichs, M.S., Bell, A.J., Ernst, E.G., Lavoie, T.B., and Mueller, L. (1993). Characterization of the three-dimensional solution structure of human profilin: 1H, 13C, and 15N NMR assignments and global folding pattern. Biochemistry *32*, 13818-13829.

Meyer, N., and Penn, L.Z. (2008). Reflecting on 25 years with MYC. Nat Rev Cancer *8*, 976-990.

Michelotti, G.A., Michelotti, E.F., Pullner, A., Duncan, R.C., Eick, D., and Levens, D. (1996). Multiple single-stranded cis elements are associated with activated chromatin of the human c-myc gene in vivo. Mol Cell Biol *16*, 2656-2669.

Min, H., Turck, C.W., Nikolic, J.M., and Black, D.L. (1997). A new regulatory protein, KSRP, mediates exon inclusion through an intronic splicing enhancer. Genes Dev *11*, 1023-1036.

Nair, S.K., and Burley, S.K. (2006). Structural aspects of interactions within the Myc/Max/Mad network. Curr Top Microbiol Immunol *302*, 123-143.

Nakagawa, M., Koyanagi, M., Tanabe, K., Takahashi, K., Ichisaka, T., Aoi, T., Okita, K., Mochiduki, Y., Takizawa, N., and Yamanaka, S. (2008). Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts. Nat Biotechnol *26*, 101-106.

Nau, M.M., Brooks, B.J., Battey, J., Sausville, E., Gazdar, A.F., Kirsch, I.R., McBride, O.W., Bertness, V., Hollis, G.F., and Minna, J.D. (1985). L-myc, a new myc-related gene amplified and expressed in human small cell lung cancer. Nature *318*, 69-73.

Nechaev, S., and Adelman, K. (2008). Promoter-proximal Pol II: when stalling speeds things up. Cell Cycle *7*, 1539-1544.

Nesbit, C.E., Tersak, J.M., and Prochownik, E.V. (1999). MYC oncogenes and human neoplastic disease. Oncogene *18*, 3004-3016.

Neuhaus, D., and Williamson, M.P. (2000). The Nuclear Overhauser Effect in Structural and Conformational Analysis (Second Edition) (New York: Wiley-VCH, Inc.).

Nietlispach, D., Mott, H.R., Stott, K.M., Nielsen, P.R., Thiru, A., and Laue, E.D. (2004). Structure determination of protein complexes by NMR. Methods Mol Biol *278*, 255-288.

Nilges, M. (1996). Structure calculation from NMR data. Curr Opin Struct Biol *6*, 617-623.

Nilges, M., Macias, M.J., O'Donoghue, S.I., and Oschkinat, H. (1997). Automated NOESY interpretation with ambiguous distance restraints: the refined NMR solution structure of the pleckstrin homology domain from beta-spectrin. J Mol Biol *269*, 408-422.

Nilges, M., and O'Donoghue, S.I. (1998). Ambiguous NOEs and automated NOE assignment. Progress in Nuclear Magnetic Resonance Spectroscopy *32*, 107-139.

Oster, S.K., Ho, C.S., Soucie, E.L., and Penn, L.Z. (2002). The myc oncogene: MarvelouslY Complex. Adv Cancer Res *84*, 81-154.

Ottmann, C., Weyand, M., Sassa, T., Inoue, T., Kato, N., Wittinghofer, A., and Oecking, C. (2009). A structural rationale for selective stabilization of anti-tumor interactions of 14-3-3 proteins by cotylenin A. J Mol Biol *386*, 913-919.

Page-McCaw, P.S., Amonlirdviman, K., and Sharp, P.A. (1999). PUF60: a novel U2AF65-related splicing activity. RNA *5*, 1548-1560.

Palmer, A.G., Cavanagh, J., Wright, P.E., and Rance, M. (1991). Sensitivity improvement in proton-detected two-dimensional heteronuclear correlation NMR spectroscopy. Journal of Magnetic Resonance (1969) *93*, 151-170.

Polshakov, V.I., Birdsall, B., and Feeney, J. (2006). Effects of co-operative ligand binding on protein amide NH hydrogen exchange. J Mol Biol *356*, 886-903.

Price, S.R., Evans, P.R., and Nagai, K. (1998). Crystal structure of the spliceosomal U2B"-U2A' protein complex bound to a fragment of U2 small nuclear RNA. Nature *394*, 645-650.

Prokipcak, R.D., Herrick, D.J., and Ross, J. (1994). Purification and properties of a protein that binds to the C-terminal coding region of human c-myc mRNA. J Biol Chem *269*, 9261-9269.

Rehbein, M., Wege, K., Buck, F., Schweizer, M., Richter, D., and Kindler, S. (2002). Molecular characterization of MARTA1, a protein interacting with the dendritic targeting element of MAP2 mRNAs. J Neurochem *82*, 1039-1046.

Rehm, T., Huber, R., and Holak, T.A. (2002). Application of NMR in structural proteomics: screening for proteins amenable to structural analysis. Structure *10*, 1613-1618.

Rideau, A.P., Gooding, C., Simpson, P.J., Monie, T.P., Lorenz, M., Huttelmaier, S., Singer, R.H., Matthews, S., Curry, S., and Smith, C.W. (2006). A peptide motif in Raver1 mediates splicing repression by interaction with the PTB RRM2 domain. Nat Struct Mol Biol *13*, 839-848.

Rothe, F., Gueydan, C., Bellefroid, E., Huez, G., and Kruys, V. (2006). Identification of FUSE-binding proteins as interacting partners of TIA proteins. Biochem Biophys Res Commun *343*, 57-68.

Schanda, P., Kupce, E., and Brutscher, B. (2005). SOFAST-HMQC experiments for recording two-dimensional heteronuclear correlation spectra of proteins within a few seconds. J Biomol NMR *33*, 199-211.

Schellenberg, M.J., Edwards, R.A., Ritchie, D.B., Kent, O.A., Golas, M.M., Stark, H., Luhrmann, R., Glover, J.N., and MacMillan, A.M. (2006). Crystal structure of a core spliceosomal protein interface. Proc Natl Acad Sci U S A *103*, 1266-1271.

Schwab, M., Alitalo, K., Klempnauer, K.H., Varmus, H.E., Bishop, J.M., Gilbert, F., Brodeur, G., Goldstein, M., and Trent, J. (1983). Amplified DNA with limited homology to myc cellular oncogene is shared by human neuroblastoma cell lines and a neuroblastoma tumour. Nature *305*, 245-248.

Selenko, P., Gregorovic, G., Sprangers, R., Stier, G., Rhani, Z., Kramer, A., and Sattler, M. (2003). Structural basis for the molecular recognition between human splicing factors U2AF65 and SF1/mBBP. Mol Cell *11*, 965-976.

Shamoo, Y., Abdul-Manan, N., and Williams, K.R. (1995). Multiple RNA binding domains (RBDs) just don't add up. Nucleic Acids Res *23*, 725-728.

Shamoo, Y., Krueger, U., Rice, L.M., Williams, K.R., and Steitz, T.A. (1997). Crystal structure of the two RNA binding domains of human hnRNP A1 at 1.75 A resolution. Nat Struct Biol *4*, 215-222.

Sheik, S.S., Ananthalakshmi, P., Bhargavi, G.R., and Sekar, K. (2003). CADB: Conformation Angles DataBase of proteins. Nucleic Acids Res *31*, 448-451.

Sheiness, D., and Bishop, J.M. (1979). DNA and RNA from uninfected vertebrate cells contain nucleotide sequences related to the putative transforming gene of avian myelocytomatosis virus. J Virol *31*, 514-521.

Sheiness, D., Fanshier, L., and Bishop, J.M. (1978). Identification of nucleotide sequences which may encode the oncogenic capacity of avian retrovirus MC29. J Virol *28*, 600-610.

Siomi, H., Matunis, M.J., Michael, W.M., and Dreyfuss, G. (1993). The pre-mRNA binding K protein contains a novel evolutionarily conserved motif. Nucleic Acids Res *21*, 1193-1198.

Soucek, L., Whitfield, J., Martins, C.P., Finch, A.J., Murphy, D.J., Sodir, N.M., Karnezis, A.N., Swigart, L.B., Nasi, S., and Evan, G.I. (2008). Modelling Myc inhibition as a cancer therapy. Nature *455*, 679-683.

Sparanese, D., and Lee, C.H. (2007). CRD-BP shields c-myc and MDR-1 RNA from endonucleolytic attack by a mammalian endoribonuclease. Nucleic Acids Res *35*, 1209-1221.

Stewart, D.E., Sarkar, A., and Wampler, J.E. (1990). Occurrence and role of cis peptide bonds in protein structures. J Mol Biol *214*, 253-260.

Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell *126*, 663-676.

Trabucchi, M., Briata, P., Garcia-Mayoral, M., Haase, A.D., Filipowicz, W., Ramos, A., Gherzi, R., and Rosenfeld, M.G. (2009). The RNA-binding protein KSRP promotes the biogenesis of a subset of microRNAs. Nature *459*, 1010-1014.

Valverde, R., Edwards, L., and Regan, L. (2008). Structure and function of KH domains. FEBS J *275*, 2712-2726.

Venters, R.A., Farmer, B.T., 2nd, Fierke, C.A., and Spicer, L.D. (1996). Characterizing the use of perdeuteration in NMR studies of large proteins: 13C, 15N and 1H assignments of human carbonic anhydrase II. J Mol Biol *264*, 1101-1116.

Vita, M., and Henriksson, M. (2006). The Myc oncoprotein as a therapeutic target for human cancer. Semin Cancer Biol *16*, 318-330.

Vitali, F., Henning, A., Oberstrass, F.C., Hargous, Y., Auweter, S.D., Erat, M., and Allain, F.H. (2006). Structure of the two most C-terminal RNA recognition motifs of PTB using segmental isotope labeling. EMBO J *25*, 150-162.

Vuister, G.W., and Bax, A. (1993). Quantitative J correlation - a new approach for measuring homonuclear 3-bond J(H(N)H(alpha) coupling-constants in N-15-enriched proteins. J. Am. Chem. Soc. *115*, 7772-7777.

Vuister, G.W., and Bax, A. (1994). Measurement of four-bond HN-H alpha J-couplings in staphylococcal nuclease. J Biomol NMR *4*, 193-200.

Wang, X., Avigan, M., and Norgren, R.B., Jr. (1998). FUSE-binding protein is developmentally regulated and is highly expressed in mouse and chicken embryonic brain. Neurosci Lett *252*, 191-194.

Weber, A., Kristiansen, I., Johannsen, M., Oelrich, B., Scholmann, K., Gunia, S., May, M., Meyer, H.A., Behnke, S., Moch, H., and Kristiansen, G. (2008). The FUSE binding proteins FBP1 and FBP3 are potential c-myc regulators in renal, but not in prostate and bladder cancer. BMC Cancer *8*, 369.

Weber, A., Liu, J., Collins, I., and Levens, D. (2005). TFIIH operates through an expanded proximal promoter to fine-tune c-myc expression. Mol Cell Biol *25*, 147-161.

Wierstra, I., and Alves, J. (2008). The c-myc promoter: still MysterY and challenge. Adv Cancer Res *99*, 113-333.

Williamson, M.P., and Craven, C.J. (2009). Automated protein structure calculation from NMR data. J Biomol NMR *43*, 131-143.

Wishart, D.S., Bigam, C.G., Holm, A., Hodges, R.S., and Sykes, B.D. (1995). 1H, 13C and 15N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects. J Biomol NMR *5*, 67-81.

Xu, R.M., Jokhan, L., Cheng, X., Mayeda, A., and Krainer, A.R. (1997). Crystal structure of human UP1, the domain of hnRNP A1 that contains two RNA-recognition motifs. Structure *5*, 559-570.

Yamamoto, A., Hen, R., and Dauer, W.T. (2001). The ons and offs of inducible transgenic technology: a review. Neurobiol Dis *8*, 923-932.

Yamazaki, T., Lee, W., Arrowsmith, C.H., Muhandiram, D.R., and Kay, L.E. (1994). A suite of triple resonance NMR experiments for the backbone assignment of 15N, 13C, 2H labeled proteins with high sensitivity. Journal of the American Chemical Society *116*, 11655-11666.

Zwahlen, C., Legault, P., Vincent, S.J.F., Greenblatt, J., Konrat, R., and Kay, L.E. (1997). Methods for measurement of intermolecular NOEs by multinuclear NMR spectroscopy: Application to a bacteriophage lambda N-peptide/boxB RNA complex. Journal of the American Chemical Society *119*, 6711-6721.

Appendices

Appendix I Amino acid sequences of FIR and FBP (NP_003893) proteins. Sequences of human (NP_055096) and rat (AF165892) FIR proteins are aligned. The region cloned in the longest FIR RRM1-RRM2 construct is highlighted in yellow and it is 100% identical between the two proteins. For both FIR and FBP proteins the boundaries of domains are indicated above.

```
FIR      1   MATATIALQV NGQQGGGSEP -----AAAAA VVAAGDKWKP
rFIR    19   MATATIALQV NGQQGGGSEP AAAAAAAAAA VVAAGDKWKP
             ********** **********      ***** **********

FIR     36   PQGTDSIKME NGQSTAAKLG LPPLTPEQQE ALQKAKKYAM
rFIR    59   PQGTESIKME NGQSTGTKLG LPPLTPEQQE ALQKAKKYAM
             ****:***** *****.:*** ********** **********

                                                    RRM1
FIR     76   EQSIKSVLVK QTIAHQQQQL TNLQMAAQRQ RALAIMCRVY
rFIR    99   EQSIKSVLVK QTIAHQQQQL TNLQMAAQRQ RALAIMCRVY
             ********** ********** ********** **********

                                RRM1
FIR    116   VGSIYYELGE DTIRQAFAPF GPIKSIDMSW DSVTMKHKGF
rFIR   139   VGSIYYELGE DTIRQAFAPF GPIKSIDMSW DSVTMKHKGF
             ********** ********** ********** **********

                                RRM1
FIR    156   AFVEYEVPEA AQLALEQMNS VMLGGRNIKV GRPSNIGQAQ
rFIR   179   AFVEYEVPEA AQLALEQMNS VMLGGRNIKV GRPSNIGQAQ
             ********** ********** ********** **********

                                          RRM2
FIR    196   PIIDQLAEEA RAFNRIYVAS VHQDLSDDDI KSVFEAFGKI
rFIR   219   PIIDQLAEEA RAFNRIYVAS VHQDLSDDDI KSVFEAFGKI
             ********** ********** ********** **********
```

```
                                RRM2
                    ___________________________________________________
FIR    236    KSCTLARDPT TGKHKGYGFI EYEKAQSSQD AVSSMNLFDL
rFIR   259    KSCTLARDPT TGKHKGYGFI EYEKAQSSQD AVSSMNLFDL
              ********** ********** ********** **********

                     RRM2
                    ___________
FIR    276    GGQYLRVGKA VTPPMPLLTP ATPGGLPPAA AVAAAAATAK
rFIR   299    GGQYLRVGKA VTPPMPLLTP ATPGGLPPAA AVAAAAATAK
              ********** ********** ********** **********

FIR    316    ITAQEAVAGA AVLGTLGTPG LVSPALTLAQ PLGTLPQAVM
rFIR   339    ITAQEAVAGA AVLGTLATPG LVSPALTLAQ PLGALPQAVM
              ********** ******.*** ********** ***:******

FIR    356    AAQAPGVITG VTPARPPIPV TIPSVGVVNP ILASPPTLGL
rFIR   379    AAQAPGVITG VTPARPPIPV TIPSVGVVNP ILASPPTLGL
              ********** ********** ********** **********

FIR    396    LEPKKEKEEE ELFPESERPE MLSEQEHMSI SGSSARHMVM
rFIR   419    LEPKKEKEEE ELFPESERPE MLSEQEHMSI SGSSARHMVM
              ********** ********** ********** **********

                                   RRM3/UHM
                    ___________________________________________
FIR    436    QKLLRKQEST VMVLRNMVDP KDIDDDLEGE VTEECGKFGA
rFIR   459    QKLLRKQEST VMVLRNMVDP KDIDDDLEGE VTEECGKFGA
              ********** ********** ********** **********

                                RRM3/UHM
                    ___________________________________________
FIR    476    VNRVIIYQEK QGEEEDAEII VKIFVEFSIA SETHKAIQAL
rFIR   499    VNRVIIYQEK QGEEEDAEII VKIFVEFSMA SETHKAIQAL
              ********** ********** ********:* **********
```

```
                RRM3/UHM
FIR    516   NGRWFAGRKV VAEVYDQERF DNSDLSA
rFIR   539   NGRWFGGRKV VAEVYDQERF DNSDLSA
              *****.**** ********** *******
```

```
                                                    Nbox
FBP      1   MADYSTVPPP SSGSAGGGGG GGGGGGVNDA FKDALQRARQ

                Nbox
FBP     41   IAAKIGGDAG TSLNSNDYGY GGQKRPLEDG DQPDAKKVAP

                                              KH1
FBP     81   QNDSFGTQLP PMHQQQSRSV MTEEYKVPDG MVGFIIGRGG

                            KH1
FBP    121   EQISRIQQES GCKIQIAPDS GGLPERSCML TGTPESVQSA

              KH1                           KH2
FBP    161   KRLLDQIVEK GRPAPGFHHG DGPGNAVQEI MIPASKAGLV

                            KH2
FBP    201   IGKGGETIKQ LQERAGVKMV MIQDGPQNTG ADKPLRITGD

                    KH2                             KH3
FBP    241   PYKVQQAKEM VLELIRDQGG FREVRNEYGS RIGGNEGIDV

                            KH3
FBP    281   PIPRFAVGIV IGRNGEMIKK IQNDAGVRIQ FKPDDGTTPE

                        KH3
FBP    321   RIAQITGPPD RCQHAAEIIT DLLRSVQAGN PGGPGPGGRG

                                        KH4
FBP    361   RGRGQGNWNM GPPGGLQEFN FIVPTGKTGL IIGKGGETIK

                            KH4
FBP    401   SISQQSGARI ELQRNPPPNA DPNMKLFTIR GTPQQIDYAR
```

KH4

FBP 441 QLIEEKIGGP VNPLGPPVPH GPHGVPGPHG PPGPPGPGTP

FBP 481 MGPYNPAPYN PGPPGPAPHG PPAPYAPQGW GNAYPHWQQQ

FBP 521 APPDPAKAGT DPNSAAWAAY YAHYYQQQAQ PPPAAPAGAP

FBP 561 TTTQTNGQGD QQNPAPAGQV DYTKAWEEYY KKMGQAVPAP

FBP 601 TGAPPGGQPD YSAAWAEYYR QQAAYYAQTS PQGMPQHPPA

FBP 641 PQGQ

Appendix II Summary of sample and experimental conditions for NMR 3D and relaxation experiments.

FIR RRM1-RRM2			
Experiment	**FIR RRM1-RRM2 conc. [mM]**	**Sample**	**^{1}H frequency of spectrometer [MHz]**
CBCA(CO)NH	0.30	^{15}N^{13}C	700
HN(CO)CACB	0.35	^{15}N^{13}C^{2}H	600
HNCA	0.30	^{15}N^{13}C	800
HNCA	0.35	^{15}N^{13}C^{2}H	700
HN(CO)CA	0.30	^{15}N^{13}C	800
HNCO	0.30	^{15}N^{13}C	800
HNCACB	0.35	^{15}N^{13}C^{2}H	700
HNCACB	0.35	^{15}N^{13}C	700
HCCH TOCSY	0.30	^{15}N^{13}C	600
^{15}N TOCSY	0.25	^{15}N	600
^{13}C NOESY HSQC	0.35	^{15}N^{13}C	800
^{13}C NOESY HSQC optimised for aromatics	0.35	^{15}N^{13}C	800
^{15}N NOESY HSQC	0.25	^{15}N	600
T1 relaxation	0.35	^{15}N	600
T2 relaxation	0.35	^{15}N	600
hetNOE	0.35	^{15}N	600

All experiments performed in 90% H_2O/10% $^{2}H_2O$ at 37 °C.

FIR RRM1-RRM2 – FBP Nbox complex						
Experiment	**FIR RRM1-RRM2 conc. [mM]**	**FBP Nbox conc. [mM]**	**Sample**	**Solvent**	**Temperature [°C]**	**^{1}H frequency of spectrometer [MHz]**
HNCA	0.30	0.60	^{15}N^{13}C protein	90% H_2O/10% $^{2}H_2O$	37	600
HCCH TOCSY	0.75	0.25	^{15}N^{13}C peptide	100% $^{2}H_2O$	37	600
HCCH TOCSY	0.60	0.20	^{15}N^{13}C peptide	100% $^{2}H_2O$	45	600
^{15}N NOESY	0.60	1.25	^{15}N protein	90% H_2O/10% $^{2}H_2O$	37	800
^{13}C-filtered ^{13}C NOESY HSQC	0.30	0.60	^{15}N^{13}C protein	100% $^{2}H_2O$	37	600
^{13}C NOESY HSQC w/o ^{13}C decoupling during the ^{1}H indirectly acquired dimension	0.30	0.60	^{15}N^{13}C protein	100% $^{2}H_2O$	37	600
^{13}C NOESY HSQC w/o ^{13}C decoupling during the ^{1}H indirectly acquired dimension	0.75	0.25	^{15}N^{13}C peptide	100% $^{2}H_2O$	37	800

FIR RRM1-RRM2 – FBP Nbox complex						
Experiment	**FIR RRM1-RRM2 conc. [mM]**	**FBP Nbox conc. [mM]**	**Sample**	**Solvent**	**Temperature [°C]**	**^{1}H frequency of spectrometer [MHz]**
^{13}C NOESY HSQC w/o ^{13}C decoupling during the ^{1}H indirectly acquired dimension	0.60	0.20	^{15}N^{13}C peptide	100% $^{2}H_2O$	45	700
^{13}C NOESY HSQC	0.55	1.45	^{15}N^{13}C protein	100% $^{2}H_2O$	37	700
^{13}C NOESY HSQC	0.55	1.45	^{15}N^{13}C protein	90% H_2O/10% $^{2}H_2O$	37	800
^{13}C NOESY HSQC	0.85	0.30	^{15}N^{13}C peptide	100% $^{2}H_2O$	37	800
^{13}C NOESY HSQC	0.65	0.35	^{15}N^{13}C peptide	100% $^{2}H_2O$	45	700
^{13}C NOESY HSQC optimised for aromatics	0.55	1.45	^{15}N^{13}C protein	100% $^{2}H_2O$	37	800
T1 relaxation	0.40	0.80	^{15}N protein	90% H_2O/10% $^{2}H_2O$	37	600
T2 relaxation	0.40	0.80	^{15}N protein	90% H_2O/10% $^{2}H_2O$	37	600
hetNOE	0.40	0.80	^{15}N protein	90% H_2O/10% $^{2}H_2O$	37	600

FBP Nbox		
Experiment	**FBP Nbox conc. [mM]**	**^{1}H frequency of spectrometer [MHz]**
HNCACB	0.40	600
CBCA(CO)NH	0.35	700
^{15}N TOCSY	0.35	600
HCCH TOCSY	0.35	700

All experiments performed on $^{15}N^{13}C$ sample in 90% H_2O/10% $^{2}H_2O$ at 37 °C.

FIR RRM1-RRM2 – ssFUSE29 complex			
Experiment	**FIR RRM1-RRM2 conc. [mM]**	**ssFUSE29 conc. [mM]**	**^{1}H frequency of spectrometer [MHz]**
T1 relaxation	0.40	0.20	600
T2 relaxation	0.40	0.20	600
hetNOE	0.40	0.20	600

All experiments performed on ^{15}N sample in 90% H_2O/10% $^{2}H_2O$ at 37 °C.

Monitored – analyte (start conc. [μM])	**Titrated in – titrant (stock conc. [mM])**	**Experiment recorded**	**Titration steps (equivalents)**	**pH**
FIR RRM1-RRM2 (41)	FBP Nbox (1.0)	sofast ^{15}N HMQC	0, 0.2, 0.4, 0.6, 0.8, 1.2, 1.6, 2, 3, 4, 6.1, 10.2	7.4
FIR RRM1-RRM2 (41)	FBP2 Nbox (1.0)	sofast ^{15}N HMQC	0, 0.13, 0.25, 0.4, 0.5, 0.75, 1, 1.25, 1.9, 2.5, 3.8, 6.3	7.4
FIR RRM1-RRM2 (41)	FBP3 Nbox (1.1)	sofast ^{15}N HMQC	0, 0.1, 0.3, 0.6, 1.15, 1.7, 2.3, 2.9, 3.5, 4.6, 5.8, 6.9, 8.6, 11.5, 14.3	7.4
FIR RRM1-RRM2 (41)	FBP Nbox with Y (2.0)	sofast ^{15}N HMQC	0, 0.1, 0.25, 0.4, 0.5, 0.75, 1, 1.2, 1.8, 2.4, 3.7, 6.1	7.4
FIR RRM1-RRM2 (41)	FBP2 Nbox with Y (2.0)	sofast ^{15}N HMQC	0, 0.1, 0.25, 0.4, 0.5, 0.75, 1, 1.2, 1.8, 2.4, 3.7, 6.1	7.4
FIR RRM1-RRM2 (75)	ssFUSE29 (1.5)	sofast ^{15}N HMQC	0, 0.1, 0.2, 0.4, 0.6, 0.8, 1, 1.5	7.4
FIR RRM1-RRM2 (25)	TTTTT (3.1)	sofast ^{15}N HMQC	0, 0.2, 0.5, 1, 2, 3, 5, 8	7.4
FIR RRM1-RRM2 (25)	TGTGT (3.0)	sofast ^{15}N HMQC	0, 0.2, 0.5, 1, 2, 3, 5, 8	7.4
FIR RRM1-RRM2 (25)	TATAT (3.0)	sofast ^{15}N HMQC	0, 0.2, 1, 3, 6, 9, 14	7.4
FIR RRM1-RRM2 (25)	AAAAA (3.0)	sofast ^{15}N HMQC	0, 0.2, 1, 3, 6, 8	7.4
FIR RRM1-RRM2 (25)[a]	TGTGT (2.5)	sofast ^{15}N HMQC	0, 0.2, 0.5, 1, 2, 3, 5, 8	7.4
FIR RRM1-RRM2 (41)[b]	FBP Nbox with Y (2.0)	sofast ^{15}N HMQC	0, 0.1, 0.25, 0.4, 0.5, 0.73, 1, 1.2, 1.8, 2.4, 3.7, 6.1	7.4

Monitored – analyte (start conc. [μM])	**Titrated in – titrant (stock conc. [mM])**	**Experiment recorded**	**Titration steps (equivalents)**	**pH**
FIR RRM1-RRM2 (70)[c]	FBP Nbox (2.0)	sofast ^{15}N HMQC	0, 6	7.4
FIR RRM1-RRM2 (25)	DNA pool (2.0 – 3.0)[d]	sofast ^{15}N HMQC	0, 1, 4	7.4
FBP Nbox (50)	FIR RRM1-RRM2 (0.70)	sofast ^{15}N HMQC	0, 0.2, 0.5, 1, 2, 3	6.5
FBP Nbox (50)	FIR RRM1-RRM2 (0.70)	^{13}C HSQC	0, 0.5, 1, 2, 3	8.0
FBP Nbox (71)	FIR RRM1-RRM2 (0.38)	sofast ^{15}N HMQC, ^{13}C HSQC	0, 0.2, 0.5, 1, 2, 3	5.0

[a] a three-molecular titration where FIR RRM1-RRM2 – FBP Nbox (with Y) complex at 1 : 6 ratio was titrated with TGTGT oligonucleotide

[b] a three-molecular titration where FIR RRM1-RRM2 – TGTGT complex at 1 : 8 ratio was titrated with FBP Nbox (with Y)

[c] a three-molecular titration where FIR RRM1-RRM2 – ssFUSE29 complex at 1 : 1.5 ratio was titrated with FBP Nbox (with Y)

[d] SIA analysis – 16 different DNA pools where used

Appendix IV SIA analysis of FIR RRM1-RRM2 protein.

Each of the 16 pools of 5mer DNA oligonucleotides used for SIA analysis has a fixed nucleotide at one position while the other positions are equimolarly occupied by A, G, C and T nucleotides. As an example, a full composition of nANNN pool is given below (N means any nucleotide). A fixed nucleotide is in blue. A lowercase indicates that this position was not examined during SIA analysis of FIR RRM1-RRM2 protein.

Oligonucleotide composition of the nANNN DNA pool							
aAAAA	**aAGAA**	**aACAA**	**aATAA**	**gAAAA**	**gAGAA**	**gACAA**	**gATAA**
aAAAG	**aAGAG**	**aACAG**	**aATAG**	**gAAAG**	**gAGAG**	**gACAG**	**gATAG**
aAAAC	**aAGAC**	**aACAC**	**aATAC**	**gAAAC**	**gAGAC**	**gACAC**	**gATAC**
aAAAT	**aAGAT**	**aACAT**	**aATAT**	**gAAAT**	**gAGAT**	**gACAT**	**gATAT**
aAAGA	**aAGGA**	**aACGA**	**aATGA**	**gAAGA**	**gAGGA**	**gACGA**	**gATGA**
aAAGG	**aAGGG**	**aACGG**	**aATGG**	**gAAGG**	**gAGGG**	**gACGG**	**gATGG**
aAAGC	**aAGGC**	**aACGC**	**aATGC**	**gAAGC**	**gAGGC**	**gACGC**	**gATGC**
aAAGT	**aAGGT**	**aACGT**	**aATGT**	**gAAGT**	**gAGGT**	**gACGT**	**gATGT**
aAACA	**aAGCA**	**aACCA**	**aATCA**	**gAACA**	**gAGCA**	**gACCA**	**gATCA**
aAACG	**aAGCG**	**aACCG**	**aATCG**	**gAACG**	**gAGCG**	**gACCG**	**gATCG**
aAACC	**aAGCC**	**aACCC**	**aATCC**	**gAACC**	**gAGCC**	**gACCC**	**gATCC**
aAACT	**aAGCT**	**aACCT**	**aATCT**	**gAACT**	**gAGCT**	**gACCT**	**gATCT**
aAATA	**aAGTA**	**aACTA**	**aATTA**	**gAATA**	**gAGTA**	**gACTA**	**gATTA**
aAATG	**aAGTG**	**aACTG**	**aATTG**	**gAATG**	**gAGTG**	**gACTG**	**gATTG**
aAATC	**aAGTC**	**aACTC**	**aATTC**	**gAATC**	**gAGTC**	**gACTC**	**gATTC**
aAATT	**aAGTT**	**aACTT**	**aATTT**	**gAATT**	**gAGTT**	**gACTT**	**gATTT**
cAAAA	**cAGAA**	**cACAA**	**cATAA**	**tAAAA**	**tAGAA**	**tACAA**	**tATAA**
cAAAG	**cAGAG**	**cACAG**	**cATAG**	**tAAAG**	**tAGAG**	**tACAG**	**tATAG**
cAAAC	**cAGAC**	**cACAC**	**cATAC**	**tAAAC**	**tAGAC**	**tACAC**	**tATAC**
cAAAT	**cAGAT**	**cACAT**	**cATAT**	**tAAAT**	**tAGAT**	**tACAT**	**tATAT**
cAAGA	**cAGGA**	**cACGA**	**cATGA**	**tAAGA**	**tAGGA**	**tACGA**	**tATGA**
cAAGG	**cAGGG**	**cACGG**	**cATGG**	**tAAGG**	**tAGGG**	**tACGG**	**tATGG**
cAAGC	**cAGGC**	**cACGC**	**cATGC**	**tAAGC**	**tAGGC**	**tACGC**	**tATGC**
cAAGT	**cAGGT**	**cACGT**	**cATGT**	**tAAGT**	**tAGGT**	**tACGT**	**tATGT**
cAACA	**cAGCA**	**cACCA**	**cATCA**	**tAACA**	**tAGCA**	**tACCA**	**tATCA**
cAACG	**cAGCG**	**cACCG**	**cATCG**	**tAACG**	**tAGCG**	**tACCG**	**tATCG**
cAACC	**cAGCC**	**cACCC**	**cATCC**	**tAACC**	**tAGCC**	**tACCC**	**tATCC**
cAACT	**cAGCT**	**cACCT**	**cATCT**	**tAACT**	**tAGCT**	**tACCT**	**tATCT**
cAATA	**cAGTA**	**cACTA**	**cATTA**	**tAATA**	**tAGTA**	**tACTA**	**tATTA**
cAATG	**cAGTG**	**cACTG**	**cATTG**	**tAATG**	**tAGTG**	**tACTG**	**tATTG**
cAATC	**cAGTC**	**cACTC**	**cATTC**	**tAATC**	**tAGTC**	**tACTC**	**tATTC**
cAATT	**cAGTT**	**cACTT**	**cATTT**	**tAATT**	**tAGTT**	**tACTT**	**tATTT**

The procedure for calculation of SIA scores and the numerical values obtained during analysis of FIR RRM1-RRM2 protein are reported below.

1. For each peak analysed (in total 15 peaks), average chemical shift perturbations $\Delta\delta_{avg}$ at the protein : DNA ratio of 1 : 4 were calculated using equation:

$$\Delta\delta_{avg} = \sqrt{\left(\frac{\Delta\delta_N}{10}\right)^2 + (\Delta\delta_H)^2}$$

   where:

   $\Delta\delta_N$ chemical shift change in ^{15}N dimension

   $\Delta\delta_H$ chemical shift change in ^{1}H dimension

   The values of $\Delta\delta_{avg}$ obtained are reported below.

2. For each peak, the values of $\Delta\delta_{avg}$ were normalised to the highest value of $\Delta\delta_{avg}$ obtained for titrations with four different DNA pools with a fixed nucleotide in one position.
3. For each DNA pool, the normalised values of $\Delta\delta_{avg}$ were averaged over all 15 peaks to yield a final SIA score for a given nucleotide at a given position.

Nucleotide preference of FIR RRM1-RRM2 for the first position tested									
		DNA pool used for the titration							
		$\Delta\delta_{avg}$ [ppm]				$\Delta\delta_{avg}$ normalised to the highest value			
		nANNN	nGNNN	nCNNN	nTNNN	nANNN	nGNNN	nCNNN	nTNNN
Peak analysed	R113	0.0241	0.0343	0.0313	0.0371	0.65	0.92	0.84	1.00
	V114	0.0137	0.0388	0.0190	0.0286	0.35	1.00	0.49	0.74
	Y115	0.0523	0.1130	0.1080	0.1510	0.35	0.75	0.72	1.00
	G117	0.0326	0.0308	0.0371	0.0958	0.34	0.32	0.39	1.00
	G124	0.0111	0.0237	0.0371	0.0601	0.18	0.39	0.62	1.00
	I128	0.0138	0.0352	0.0231	0.0431	0.32	0.82	0.54	1.00
	A133	0.0232	0.0240	0.0134	0.0049	0.97	1.00	0.56	0.20
	D142	0.0254	0.0406	0.0250	0.0407	0.62	1.00	0.61	1.00
	W145	0.0447	0.0696	0.0570	0.0696	0.64	1.00	0.82	1.00
	D146	0.0229	0.0546	0.0396	0.0532	0.42	1.00	0.73	0.97
	V158	0.0123	0.0396	0.0192	0.0216	0.31	1.00	0.48	0.55
	N174	0.0360	0.0320	0.0190	0.0140	1.00	0.89	0.53	0.39
	N182	0.0198	0.0382	0.0396	0.0583	0.34	0.66	0.68	1.00
	I183	0.0160	0.0274	0.0198	0.0283	0.57	0.97	0.70	1.00
	V185	0.0295	0.0556	0.0740	0.1216	0.24	0.46	0.61	1.00
Final SIA score for a given nucleotide in the first position tested						0.49	0.81	0.62	0.86

Nucleotide preference of FIR RRM1-RRM2 for the second position tested									
		DNA pool used for the titration							
		$\Delta\delta_{avg}$ [ppm]				$\Delta\delta_{avg}$ normalised to the highest value			
		nNANN	nNGNN	nNCNN	nNTNN	nNANN	nNGNN	nNCNN	nNTNN
Peak analysed	R113	0.0101	0.0400	0.0239	0.0410	0.25	0.98	0.58	1.00
	V114	0.0185	0.0342	0.0219	0.0229	0.54	1.00	0.64	0.67
	Y115	0.0750	0.1279	0.0040	0.1476	0.51	0.87	0.03	1.00
	G117	0.0237	0.0330	0.0450	0.0529	0.45	0.62	0.85	1.00
	G124	0.0053	0.0336	0.0311	0.0388	0.14	0.87	0.80	1.00
	I128	0.0180	0.0220	0.0231	0.0360	0.50	0.61	0.64	1.00
	D142	0.0087	0.0351	0.0332	0.0423	0.21	0.83	0.78	1.00
	W145	0.0585	0.0838	0.0787	0.0867	0.67	0.97	0.91	1.00
	D146	0.0134	0.0498	0.0454	0.0411	0.27	1.00	0.91	0.83
	A156	0.0385	0.0838	0.0943	0.1045	0.37	0.80	0.90	1.00
	F157	0.0659	0.1156	0.0208	0.1205	0.55	0.96	0.17	1.00
	N174	0.0052	0.0361	0.0251	0.0184	0.14	1.00	0.70	0.51
	N182	0.0158	0.0381	0.0430	0.0464	0.34	0.82	0.93	1.00
	I183	0.0094	0.0272	0.0129	0.0247	0.35	1.00	0.47	0.91
	V185	0.2242	0.0447	0.0831	0.1025	1.00	0.20	0.37	0.46
Final SIA score for a given nucleotide in the second position tested						0.42	0.84	0.65	0.89

Nucleotide preference of FIR RRM1-RRM2 for the third position tested									
		DNA pool used for the titration							
		$\Delta\delta_{avg}$ [ppm]				$\Delta\delta_{avg}$ normalised to the highest value			
		nNNAN	nNNGN	nNNCN	nNNTN	nNNAN	nNNGN	nNNCN	nNNTN
Peak analysed	R113	0.0158	0.0125	0.0285	0.0464	0.34	0.27	0.61	1.00
	V114	0.0098	0.0177	0.0110	0.0499	0.20	0.35	0.22	1.00
	Y115	0.0383	0.0417	0.0603	0.0088	0.64	0.69	1.00	0.15
	G117	0.0203	0.0235	0.0325	0.0540	0.38	0.44	0.60	1.00
	G124	0.0079	0.0150	0.0205	0.0383	0.21	0.39	0.54	1.00
	I128	0.0102	0.0087	0.0164	0.0277	0.37	0.31	0.59	1.00
	D142	0.0090	0.0427	0.0271	0.0315	0.21	1.00	0.63	0.74
	W145	0.0326	0.0428	0.0788	0.0815	0.40	0.53	0.97	1.00
	D146	0.0132	0.0255	0.0393	0.0591	0.22	0.43	0.66	1.00
	F157	0.0372	0.0720	0.0828	0.1276	0.29	0.56	0.65	1.00
	V158	0.0137	0.0080	0.0109	0.0215	0.64	0.37	0.51	1.00
	N174	0.0074	0.0122	0.0121	0.0196	0.38	0.62	0.62	1.00
	N182	0.0176	0.0407	0.0498	0.0750	0.23	0.54	0.66	1.00
	I183	0.0115	0.0213	0.0190	0.0373	0.31	0.57	0.51	1.00
	V185	0.0375	0.0301	0.0604	0.1383	0.27	0.22	0.44	1.00
Final SIA score for a given nucleotide in the third position tested						0.34	0.49	0.61	0.93

Nucleotide preference of FIR RRM1-RRM2 for the fourth position tested									
		DNA pool used for the titration							
		$\Delta\delta_{avg}$ [ppm]				$\Delta\delta_{avg}$ normalised to the highest value			
		nNNNA	nNNNG	nNNNC	nNNNT	nNNNA	nNNNG	nNNNC	nNNNT
Peak analysed	R113	0.0335	0.0350	0.0370	0.0415	0.81	0.84	0.89	1.00
	V114	0.0207	0.0358	0.0246	0.0618	0.33	0.58	0.40	1.00
	G117	0.0225	0.0284	0.0505	0.0450	0.45	0.56	1.00	0.89
	G124	0.0219	0.0263	0.0252	0.0241	0.83	1.00	0.96	0.92
	I128	0.0165	0.0199	0.0134	0.0372	0.44	0.53	0.36	1.00
	D142	0.0276	0.0311	0.0205	0.0278	0.89	1.00	0.74	0.89
	W145	0.0667	0.0935	0.0623	0.1277	0.52	0.73	0.49	1.00
	D146	0.0449	0.0589	0.0435	0.0681	0.66	0.86	0.64	1.00
	A156	0.0663	0.0384	0.0627	0.0948	0.70	0.41	0.66	1.00
	F157	0.1006	0.1152	0.0675	0.1577	0.64	0.73	0.43	1.00
	V158	0.0150	0.0159	0.0101	0.0328	0.46	0.48	0.31	1.00
	N174	0.0128	0.0181	0.0082	0.0259	0.49	0.70	0.32	1.00
	N182	0.0481	0.0695	0.0481	0.0730	0.66	0.95	0.66	1.00
	I183	0.0195	0.0363	0.0189	0.0312	0.54	1.00	0.52	0.86
	V185	0.0725	0.0783	0.0716	0.1041	0.70	0.75	0.69	1.00
Final SIA score for a given nucleotide in the fourth position tested						0.61	0.74	0.60	0.97

Appendix V Backbone and side chain assignment of FIR RRM1-RRM2 and FBP Nbox in the free form and in the complex. Ambiguity value of 1 indicates unique assignments including isolated methyl protons, geminal atoms, and geminal methyl groups with identical chemical shifts (e.g. ILE HD11, HD12, HD13 protons). Ambiguity value of 2 indicates ambiguity of geminal atoms or geminal methyl proton groups (e.g. ASP HB2 and HB3 protons, LEU CD1 and CD2 carbons, or LEU HD11, HD12, HD13 and HD21, HD22, HD23 methyl protons), which were not stereospecifically assigned. Ambiguity value of 3 indicates aromatic atoms on opposite sides of symmetrical rings (e.g. TYR HE1 and HE2 protons).

Residue name	Residue number	Atom name	Chemical shift [ppm]	Ambiguity	Chemical shift [ppm]	Ambiguity
			FIR RRM1-RRM2 free		FIR RRM1-RRM2 bound	
ARG	104	CA	60.09	1		
ARG	104	CB	30.30	1		
ARG	104	C	177.29	1		
GLN	105	N	118.16	1	118.00	1
GLN	105	HN	8.25	1	8.19	1
GLN	105	CA	59.48	1	59.25	1
GLN	105	CB	28.45	1		
GLN	105	C	178.83	1		
ARG	106	N	119.39	1	119.33	1
ARG	106	HN	7.93	1	7.95	1
ARG	106	CA	59.11	1	58.99	1
ARG	106	CB	30.05	1		
ARG	106	C	178.39	1		
ALA	107	N	121.41	1	121.41	1
ALA	107	HN	7.82	1	7.82	1
ALA	107	CA	55.25	1	55.12	1
ALA	107	HA	3.94	1	3.93	1
ALA	107	CB	18.61	1	18.48	1
ALA	107	HB	1.47	1	1.46	1
ALA	107	C	178.67	1		
LEU	108	N	117.79	1	117.73	1
LEU	108	HN	8.16	1	8.14	1
LEU	108	CA	57.99	1	57.93	1
LEU	108	HA	4.01	1	3.99	1
LEU	108	CB	42.01	1	41.95	1
LEU	108	HB2/HB3	1.81/1.48	2	1.80/1.47	2
LEU	108	CG	27.43	1	27.43	1
LEU	108	HG	1.72	1	1.71	1
LEU	108	CD1/CD2	25.73/24.64	2	25.63/24.52	2
LEU	108	HD1/HD2	0.84/0.87	2	0.84/0.85	2
LEU	108	C	178.48	1		
ALA	109	N	119.35	1	119.34	1
ALA	109	HN	7.42	1	7.41	1
ALA	109	CA	55.44	1	55.21	1
ALA	109	HA	4.11	1	4.10	1
ALA	109	CB	18.07	1	18.10	1
ALA	109	HB	1.45	1	1.44	1
ALA	109	C	181.10	1		
ILE	110	N	118.48	1	118.52	1
ILE	110	HN	7.59	1	7.59	1
ILE	110	CA	65.02	1	64.97	1
ILE	110	HA	3.68	1	3.67	1
ILE	110	CB	38.39	1	38.25	1
ILE	110	HB	1.88	1	1.87	1
ILE	110	CG1	29.57	1	29.48	1
ILE	110	HG12/HG13	1.71/1.04	2	1.71/1.04	2
ILE	110	CG2	17.64	1	17.58	1
ILE	110	HG2	0.85	1	0.84	1
ILE	110	CD1	13.90	1	13.90	1
ILE	110	HD1	0.74	1	0.74	1
ILE	110	C	178.72	1		
MET	111	N	118.37	1	118.27	1
MET	111	HN	8.20	1	8.18	1
MET	111	CA	58.67	1	58.49	1

Residue name	Residue number	Atom name	Chemical shift [ppm]	Ambiguity	Chemical shift [ppm]	Ambiguity
			FIR RRM1-RRM2 free		FIR RRM1-RRM2 bound	
MET	111	HA	3.90	1	3.89	1
MET	111	CB	34.77	1	34.70	1
MET	111	HB2/HB3	2.29	2	2.32	2
MET	111	CG	32.87	1	32.87	1
MET	111	HG2/HG3	2.61	2	2.60	2
MET	111	CE	16.74	1	16.75	1
MET	111	HE	1.97	1	1.97	1
MET	111	C	176.36	1		
CYS	112	N	113.70	1	113.68	1
CYS	112	HN	7.24	1	7.23	1
CYS	112	CA	62.42	1	62.51	1
CYS	112	HA	4.16	1	4.15	1
CYS	112	CB	28.42	1	28.22	1
CYS	112	HB2/HB3	3.58/3.07	2	3.57/3.06	2
CYS	112	C	173.38	1		
ARG	113	N	120.14	1	120.10	1
ARG	113	HN	7.43	1	7.41	1
ARG	113	CA	55.81	1	55.62	1
ARG	113	HA	5.37	1	5.36	1
ARG	113	CB	33.83	1	33.72	1
ARG	113	HB2/HB3	1.81/1.74	2	1.80/1.74	2
ARG	113	CG	28.57	1	28.56	1
ARG	113	HG2/HG3	1.40/1.74	2	1.40/1.74	2
ARG	113	CD	43.84	1	43.61	1
ARG	113	HD2/HD3	2.91/3.08	2	2.91/3.07	2
ARG	113	C	174.96	1		
VAL	114	N	117.47	1	117.31	1
VAL	114	HN	9.09	1	9.08	1
VAL	114	CA	60.40	1	60.26	1
VAL	114	HA	4.72	1	4.72	1
VAL	114	CB	34.99	1	34.97	1
VAL	114	HB	2.09	1	2.09	1
VAL	114	CG1/CG2	20.53/22.90	2	20.41/22.87	2
VAL	114	HG1/HG2	0.90/0.93	2	0.89/0.92	2
VAL	114	C	174.52	1		
TYR	115	N	125.38	1	125.28	1
TYR	115	HN	8.83	1	8.82	1
TYR	115	CA	56.85	1	56.75	1
TYR	115	HA	4.49	1	4.48	1
TYR	115	CB	39.85	1	39.88	1
TYR	115	HB2/HB3	2.48/2.39	2	2.46/2.38	2
TYR	115	CD1/CD2	133.01	3	133.01	3
TYR	115	HD1/HD2	6.39	3	6.39	3
TYR	115	CE1/CE2	117.68	3	117.62	3
TYR	115	HE1/HE2	6.23	3	6.23	3
TYR	115	C	174.52	1		
VAL	116	N	127.91	1	127.88	1
VAL	116	HN	8.39	1	8.38	1
VAL	116	CA	60.38	1	60.27	1
VAL	116	HA	4.59	1	4.58	1
VAL	116	CB	33.27	1	33.26	1
VAL	116	HB	1.57	1	1.56	1
VAL	116	CG1/CG2	21.46/22.10	2	21.39/21.95	2
VAL	116	HG1/HG2	0.29/0.72	2	0.29/0.72	2
VAL	116	C	173.93	1		
GLY	117	N	111.87	1	111.88	1
GLY	117	HN	9.10	1	9.11	1
GLY	117	CA	43.33	1	43.18	1
GLY	117	HA2/HA3	3.53/4.79	2	3.53/4.78	2
SER	118	N			111.41	1
SER	118	HN			8.32	1
SER	118	CA	58.76	1	58.53	1
SER	118	HA			3.98	1
SER	118	CB	63.13	1	62.97	1
SER	118	HB2/HB3			4.00/4.35	2
ILE	119	N	117.30	1	117.27	1
ILE	119	HN	8.12	1	8.13	1
ILE	119	CA	60.56	1	60.38	1
ILE	119	HA	3.85	1	3.85	1
ILE	119	CB	38.77	1	38.68	1
ILE	119	HB	1.14	1	1.13	1
ILE	119	CG1	27.63	1	27.55	1
ILE	119	HG12/HG13	1.21/0.43	2	1.22/0.44	2
ILE	119	CG2	18.32	1	18.22	1
ILE	119	HG2	0.70	1	0.69	1
ILE	119	CD1	14.20	1	14.22	1
ILE	119	HD1	0.25	1	0.24	1
TYR	120	CA			58.45	1

Residue name	Residue number	Atom name	Chemical shift [ppm]	Ambiguity	Chemical shift [ppm]	Ambiguity
			FIR RRM1-RRM2 free		FIR RRM1-RRM2 bound	
TYR	120	HA	4.43	1	4.43	1
TYR	120	CB	38.69	1	38.64	1
TYR	120	HB2/HB3	2.76/3.15	2	2.75/3.14	2
TYR	120	CD1/CD2	132.23	3	132.24	3
TYR	120	HD1/HD2	7.09	3	7.08	3
TYR	120	CE1/CE2	119.56	3	119.47	3
TYR	120	HE1/HE2	6.85	3	6.85	3
TYR	121	CA			60.57	1
TYR	121	HA	3.66	1	3.66	1
TYR	121	CB	37.73	1	37.52	1
TYR	121	HB2/HB3	2.64/2.81	2	2.64/2.80	2
TYR	121	CD1/CD2	132.92	3	132.96	3
TYR	121	HD1/HD2	6.80	3	6.79	3
TYR	121	CE1/CE2			118.62	3
TYR	121	HE1/HE2			6.78	3
TYR	121	C	175.86	1		
GLU	122	N	128.33	1	128.30	1
GLU	122	HN	7.66	1	7.70	1
GLU	122	CA	56.53	1	56.42	1
GLU	122	HA	4.14	1	4.13	1
GLU	122	CB	30.39	1	30.34	1
GLU	122	HB2/HB3	1.57/2.16	2	1.55/2.15	2
GLU	122	CG	37.01	1	36.85	1
GLU	122	HG2/HG3	1.84/1.74	2	1.82/1.73	2
GLU	122	C	176.57	1		
LEU	123	N	120.10	1	120.00	1
LEU	123	HN	7.77	1	7.77	1
LEU	123	CA	55.02	1	54.80	1
LEU	123	HA	4.40	1	4.39	1
LEU	123	CB	42.40	1	42.42	1
LEU	123	HB2/HB3	1.37/1.93	2	1.37/1.92	2
LEU	123	CG	27.14	1	27.13	1
LEU	123	HG	1.83	1	1.82	1
LEU	123	CD1/CD2	23.06/26.27	2	23.06/26.23	2
LEU	123	HD1/HD2	0.94/0.94	2	0.93/0.94	2
LEU	123	C	176.32	1		
GLY	124	N	107.83	1	107.83	1
GLY	124	HN	8.27	1	8.26	1
GLY	124	CA	43.57	1	43.65	1
GLY	124	HA2/HA3	3.93/4.53	2	3.92/4.51	2
GLY	124	C	174.30	1		
GLU	125	N	120.08	1	120.11	1
GLU	125	HN	8.91	1	8.90	1
GLU	125	CA	60.58	1	60.43	1
GLU	125	HA	3.72	1	3.71	1
GLU	125	CB	30.25	1	30.21	1
GLU	125	HB2/HB3	2.22/2.16	2	2.21/2.16	2
GLU	125	CG	36.80	1	36.73	1
GLU	125	HG2/HG3	2.23	2	2.22	2
GLU	125	C	177.83			
ASP	126	N	116.93	1	116.89	1
ASP	126	HN	8.83	1	8.85	1
ASP	126	CA	57.47	1	57.22	1
ASP	126	HA			4.35	1
ASP	126	CB	39.81	1		
ASP	126	C	178.72	1		
THR	127	N	117.49	1	117.51	1
THR	127	HN	7.64	1	7.63	1
THR	127	CA	66.69	1	66.53	1
THR	127	HA	4.01	1	3.99	1
THR	127	CB	68.52	1	68.42	1
THR	127	HB	4.18	1	4.16	1
THR	127	CG2	22.87	1	22.78	1
THR	127	HG2	1.33	1	1.32	1
THR	127	C	176.69	1		
ILE	128	N	121.42	1	121.35	1
ILE	128	HN	7.44	1	7.43	1
ILE	128	CA	63.68	1	63.62	1
ILE	128	HA	3.73	1	3.72	1
ILE	128	CB	36.25	1	36.24	1
ILE	128	HB	2.01	1	2.00	1
ILE	128	CG1	27.68	1	27.62	1
ILE	128	HG12/HG13	1.21	2	1.20	2
ILE	128	CG2	17.61	1	17.55	1
ILE	128	HG2	0.54	1	0.53	1
ILE	128	CD1	10.80	1	10.69	1
ILE	128	HD1	0.50	1	0.50	1
ILE	128	C	177.39	1		

Residue name	Residue number	Atom name	Chemical shift [ppm]	Ambiguity	Chemical shift [ppm]	Ambiguity
			FIR RRM1-RRM2 free		FIR RRM1-RRM2 bound	
ARG	129	N	120.70	1	120.65	1
ARG	129	HN	8.89	1	8.88	1
ARG	129	CA	60.93	1	60.94	1
ARG	129	HA	3.67	1	3.67	1
ARG	129	CB	30.20	1	30.18	1
ARG	129	HB2/HB3	1.90/1.96	2	1.89/1.94	2
ARG	129	CG	28.36	1	28.33	1
ARG	129	HG2/HG3	1.54	2	1.53	2
ARG	129	CD	42.95	1	42.79	1
ARG	129	HD2/HD3	3.16/3.34	2	3.16/3.33	2
ARG	129	C	177.63	1		
GLN	130	N	115.04	1	115.03	1
GLN	130	HN	7.94	1	7.93	1
GLN	130	CA	58.93	1	58.85	1
GLN	130	HA	4.04	1	4.04	1
GLN	130	CB	28.61	1	28.59	1
GLN	130	HB2/HB3	2.18	2	2.17	2
GLN	130	CG	34.01	1	33.93	1
GLN	130	HG2/HG3	2.42/2.55	2	2.41/2.55	2
GLN	130	C	178.42	1		
ALA	131	N	119.03	1	118.99	1
ALA	131	HN	7.74	1	7.74	1
ALA	131	CA	53.89	1	53.82	1
ALA	131	HA	4.20	1	4.20	1
ALA	131	CB	19.12	1	19.16	1
ALA	131	HB	1.45	1	1.44	1
ALA	131	C	177.76	1		
PHE	132	N	112.01	1	112.04	1
PHE	132	HN	8.09	1	8.09	1
PHE	132	CA	60.29	1	60.19	1
PHE	132	HA	4.74	1	4.74	1
PHE	132	CB	38.23	1	38.12	1
PHE	132	HB2/HB3	2.85/3.54	2	2.85/3.53	2
PHE	132	CD1/CD2	132.30	3	132.34	3
PHE	132	HD1/HD2	7.96	3	7.96	3
PHE	132	CE1/CE2	129.69	3	129.65	3
PHE	132	HE1/HE2	6.88	3	6.88	3
PHE	132	C	177.63	1		
ALA	133	N	130.20	1	130.14	1
ALA	133	HN	8.65	1	8.64	1
ALA	133	CA	55.76	1	55.38	1
ALA	133	HA	4.71	1	4.71	1
ALA	133	CB	17.39	1	17.39	1
ALA	133	HB	1.62	1	1.61	1
PRO	134	CA	65.65	1	65.71	1
PRO	134	HA	4.04	1	4.03	1
PRO	134	CB	31.13	1	31.05	1
PRO	134	HB2/HB3	0.27/2.02	2	0.29/2.04	2
PRO	134	CG	28.12	1	28.24	1
PRO	134	HG2/HG3	1.71/1.78	2	1.68/1.77	2
PRO	134	CD	52.29	1	52.16	1
PRO	134	HD2/HD3	3.72/3.08	2	3.71/3.09	2
PRO	134	C	176.45	1		
PHE	135	N	108.37	1	108.22	1
PHE	135	HN	6.54	1	6.53	1
PHE	135	CA	60.00	1	60.00	1
PHE	135	HA	3.90	1	3.88	1
PHE	135	CB	38.54	1	38.47	1
PHE	135	HB2/HB3	3.47/2.62	2	3.46/2.62	2
PHE	135	CD1/CD2	131.02	3	131.04	3
PHE	135	HD1/HD2	7.27	3	7.29	3
PHE	135	CE1/CE2	131.66	3	131.65	3
PHE	135	HE1/HE2	7.37	3	7.35	3
PHE	135	CZ	129.29	1	129.53	1
PHE	135	HZ	7.13	1	7.12	1
PHE	135	C	174.19	1		
GLY	136	N	105.86	1	105.91	1
GLY	136	HN	7.72	1	7.71	1
GLY	136	CA	44.78	1	44.69	1
GLY	136	HA2/HA3	3.95/4.11	2	3.93/4.08	2
PRO	137	CA	62.67	1	62.62	1
PRO	137	HA	4.66	1	4.65	1
PRO	137	CB	31.67	1	31.66	1
PRO	137	HB2/HB3	2.36/2.06	2	2.34/2.05	2
PRO	137	CG	29.00	1	28.93	1
PRO	137	HG2/HG3	2.29/2.23	2	2.29/2.21	2
PRO	137	CD	49.62	1	49.63	1
PRO	137	HD2/HD3	3.49/3.22	2	3.47/3.18	2

Residue name	Residue number	Atom name	Chemical shift [ppm]	Ambiguity	Chemical shift [ppm]	Ambiguity
			FIR RRM1-RRM2 free		FIR RRM1-RRM2 bound	
PRO	137	C	177.08	1		
ILE	138	N	127.37	1	127.19	1
ILE	138	HN	8.79	1	8.76	1
ILE	138	CA	62.41	1	62.15	1
ILE	138	HA	3.76	1	3.75	1
ILE	138	CB	39.68	1	39.55	1
ILE	138	HB	1.65	1	1.64	1
ILE	138	CG1	29.05	1	29.00	1
ILE	138	HG12/HG13	0.33/1.92	2	0.32/1.90	2
ILE	138	CG2	17.65	1	17.60	1
ILE	138	HG2	0.54	1	0.53	1
ILE	138	CD1	14.20	1	14.22	1
ILE	138	HD1	0.77	1	0.76	1
ILE	138	C	176.34	1		
LYS	139	N	131.39	1	131.38	1
LYS	139	HN	9.53	1	9.51	1
LYS	139	CA	57.73	1	57.83	1
LYS	139	HA	4.32	1	4.31	1
LYS	139	CB	34.28	1	34.26	1
LYS	139	HB2/HB3	1.41/1.53	2	1.41/1.53	2
LYS	139	CG	24.70	1	24.64	1
LYS	139	HG2/HG3	1.30/1.38	2	1.29/1.38	2
LYS	139	HE2/HE3	2.95	2	2.94	2
LYS	139	C	176.52	1		
SER	140	N	111.00	1	110.97	1
SER	140	HN	7.73	1	7.71	1
SER	140	CA	57.83	1	57.69	1
SER	140	HA	4.49	1	4.48	1
SER	140	CB	65.38	1	65.37	1
SER	140	HB2/HB3	3.76/3.69	2	3.74/3.69	2
SER	140	C	178.45	1		
ILE	141	N	121.23	1	121.22	1
ILE	141	HN	8.20	1	8.20	1
ILE	141	CA	60.69	1	60.39	1
ILE	141	HA	4.60	1	4.58	1
ILE	141	CB	40.32	1	40.30	1
ILE	141	HB	1.64	1	1.63	1
ILE	141	CG1	28.31	1	28.23	1
ILE	141	HG12/HG13	0.84/1.50	2	0.83/1.50	2
ILE	141	CG2	16.96	1	17.03	1
ILE	141	HG2	0.60	1	0.59	1
ILE	141	CD1	13.75	1	13.72	1
ILE	141	HD1	0.76	1	0.74	1
ILE	141	C	174.15	1		
ASP	142	N	128.08	1	128.01	1
ASP	142	HN	8.92	1	8.91	1
ASP	142	CA	52.96	1	52.83	1
ASP	142	HA	5.03	1	5.02	1
ASP	142	CB	42.96	1	42.90	1
ASP	142	HB2/HB3	2.83/2.62	2	2.82/2.61	2
ASP	142	C	175.44	1		
MET	143	N	123.86	1	123.85	1
MET	143	HN	8.60	1	8.60	1
MET	143	CA	54.57	1	54.33	1
MET	143	HA	4.41	1	4.40	1
MET	143	CB	34.35	1	34.25	1
MET	143	HB2/HB3	1.66	2	1.65	2
MET	143	CG	32.19	1	32.15	1
MET	143	HG2/HG3	2.33/2.24	2	2.33/2.23	2
MET	143	CE	15.99	1	15.98	1
MET	143	HE	1.28	1	1.27	1
SER	144	CA	57.58	1	57.46	1
SER	144	HA	4.49	1	4.47	1
SER	144	CB	63.79	1	63.56	1
SER	144	HB2/HB3	3.20/3.48	2	3.17/3.46	2
SER	144	C	173.37			
TRP	145	N	122.38	1	122.33	1
TRP	145	HN	8.26	1	8.25	1
TRP	145	CA	57.50	1	57.31	1
TRP	145	HA	4.61	1	4.60	1
TRP	145	CB	31.57	1	31.52	1
TRP	145	HB2/HB3	3.07/3.01	2	3.04/3.02	2
TRP	145	CD1	127.48	1	127.28	1
TRP	145	HD1	7.13	1	7.12	1
TRP	145	NE1	129.44	1	129.40	1
TRP	145	HE1	10.12	1	10.11	1
TRP	145	CE3	120.50	1	120.54	1
TRP	145	HE3	7.34	1	7.33	1

Residue name	Residue number	Atom name	Chemical shift [ppm]	Ambiguity	Chemical shift [ppm]	Ambiguity
			FIR RRM1-RRM2 free		FIR RRM1-RRM2 bound	
TRP	145	CZ2	114.50	1	114.45	1
TRP	145	HZ2	7.45	1	7.44	1
TRP	145	CZ3	122.13	1	122.14	1
TRP	145	HZ3	7.10	1	7.09	1
TRP	145	CH2	124.64	1	124.64	1
TRP	145	HH2	7.20		7.19	1
TRP	145	C	175.26	1		
ASP	146	N	125.73	1	125.73	1
ASP	146	HN	8.28	1	8.26	1
ASP	146	CA	53.24	1	53.04	1
ASP	146	HA	4.71	1	4.73	1
ASP	146	CB	43.04	1	42.97	1
ASP	146	HB2/HB3	3.10/2.31	2	3.09/2.30	2
SER	147	CA	60.52	1	60.49	1
SER	147	CB	63.44	1		
SER	147	C	174.56	1		
VAL	148	N	121.37	1	121.40	1
VAL	148	HN	8.33	1	8.32	1
VAL	148	CA	65.25	1	65.02	1
VAL	148	HA	3.90	1	3.88	1
VAL	148	CB	31.87	1	31.83	1
VAL	148	HB	2.25	1	2.24	1
VAL	148	CG1/CG2	22.11/21.02	2	22.05/20.91	2
VAL	148	HG1/HG2	0.96/0.90	2	0.95/0.88	2
VAL	148	C	177.75	1		
THR	149	N	109.04	1	109.11	1
THR	149	HN	7.73	1	7.72	1
THR	149	CA	62.12	1	62.11	1
THR	149	HA	4.10	1	4.09	1
THR	149	CB	69.89	1	69.66	1
THR	149	HB	4.12	1	4.10	1
THR	149	CG2	21.94	1	21.80	1
THR	149	HG2	1.11	1	1.10	1
THR	149	C	175.31	1		
MET	150	N	116.67	1	116.74	1
MET	150	HN	8.06	1	8.07	1
MET	150	CA	56.15	1	55.96	1
MET	150	HA	3.90	1	3.89	1
MET	150	CB	28.98	1		
MET	150	HB2/HB3	2.20/2.38	2	2.21/2.36	2
MET	150	CE	16.64	1	16.73	1
MET	150	HE	2.03	1	2.01	1
MET	150	C	174.57	1		
LYS	151	N	117.18	1	117.12	1
LYS	151	HN	7.30	1	7.32	1
LYS	151	CA	54.72	1	54.47	1
LYS	151	HA	4.61	1	4.58	1
LYS	151	CB	35.76	1	35.64	1
LYS	151	HB2/HB3	1.40/1.65	2	1.39/1.65	2
LYS	151	CG	24.76	1	24.63	1
LYS	151	HG2/HG3	1.22	2	1.22	2
LYS	151	CD	28.97	1	28.85	1
LYS	151	HD2/HD3	1.39/1.31	2	1.38/1.30	2
LYS	151	CE	42.29	1	42.07	1
LYS	151	HE2/HE3	2.75	2	2.74	2
LYS	151	C	175.75	1		
HIS	152	N	118.08	1	117.98	1
HIS	152	HN	7.74	1	7.78	1
HIS	152	CA	54.89	1	54.74	1
HIS	152	HA	5.01	1	5.00	1
HIS	152	CB	31.02	1	31.04	1
HIS	152	HB2/HB3	2.92	2	2.91	2
HIS	152	CD2	120.34	1	120.09	1
HIS	152	HD2	6.87	1	6.87	1
HIS	152	CE1	138.47	1	138.35	1
HIS	152	HE1	7.28	1	7.28	1
PHE	155	CA	54.88	1	54.81	1
PHE	155	HA	5.52	1	5.53	1
PHE	155	CB	42.31	1	42.11	1
PHE	155	HB2/HB3	2.96/2.77	2	2.96/2.75	2
PHE	155	CD1/CD2	132.71	3	132.67	3
PHE	155	HD1/HD2	6.83	3	6.81	3
PHE	155	CE1/CE2	131.54	3	131.57	3
PHE	155	HE1/HE2	7.37	3	7.37	3
PHE	155	C	173.22	1		
ALA	156	N	122.20	1	122.21	1
ALA	156	HN	9.09	1	9.09	1
ALA	156	CA	50.34	1	50.39	1

Residue name	Residue number	Atom name	Chemical shift [ppm]	Ambiguity	Chemical shift [ppm]	Ambiguity
			FIR RRM1-RRM2 free		FIR RRM1-RRM2 bound	
ALA	156	HA	4.78	1	4.77	1
ALA	156	CB	23.38	1	23.37	1
ALA	156	HB	0.84	1	0.84	1
ALA	156	C	174.81	1		
PHE	157	N	115.41	1	115.45	1
PHE	157	HN	8.45	1	8.45	1
PHE	157	CA	56.77	1	56.66	1
PHE	157	HA	5.47	1	5.45	1
PHE	157	CB	42.15	1	42.13	1
PHE	157	HB2/HB3	2.93/2.74	2	2.92/2.75	2
PHE	157	CD1/CD2	131.83	3	131.80	3
PHE	157	HD1/HD2	7.20	3	7.20	3
PHE	157	CE1/CE2	131.45	3	131.61	3
PHE	157	HE1/HE2	7.38	3	7.38	3
PHE	157	C	175.45	1		
VAL	158	N	124.31	1	124.33	1
VAL	158	HN	8.90	1	8.90	1
VAL	158	CA	61.68	1	61.64	1
VAL	158	HA	4.09	1	4.09	1
VAL	158	CB	33.50	1	33.44	1
VAL	158	HB	1.44	1	1.43	1
VAL	158	CG1/CG2	20.89/20.29	2	20.77/20.20	2
VAL	158	HG1/HG2	0.13/0.04	2	0.13/0.04	2
VAL	158	C	173.34	1		
GLU	159	N	126.97	1	126.89	1
GLU	159	HN	8.66	1	8.64	1
GLU	159	CA	54.24	1	54.23	1
GLU	159	HA	5.01	1	5.01	1
GLU	159	CB	32.03	1	32.05	1
GLU	159	HB2/HB3	1.69/2.04	2	1.68/2.03	2
GLU	159	CG	36.77	1	36.72	1
GLU	159	HG2/HG3	2.03	2	2.02	2
GLU	159	C	175.75	1		
TYR	160	N	125.60	1	125.58	1
TYR	160	HN	8.94	1	8.94	1
TYR	160	CA	58.67	1	58.57	1
TYR	160	HA	5.11	1	5.11	1
TYR	160	CB	42.94	1	42.83	1
TYR	160	HB2/HB3	3.13/2.77	2	3.12/2.76	2
TYR	160	CD1/CD2	132.69	3	132.63	3
TYR	160	HD1/HD2	6.81	3	6.81	3
TYR	160	HE1/HE2			6.35	3
TYR	160	HH	9.40	1	9.37	1
TYR	160	C	175.86	1		
GLU	161	N	114.56	1	114.54	1
GLU	161	HN	8.71	1	8.74	1
GLU	161	CA	58.37	1	58.16	1
GLU	161	HA	4.21	1	4.20	1
GLU	161	CB	31.90	1	31.83	1
GLU	161	HB2/HB3	1.98/2.24	2	1.98/2.22	2
GLU	161	CG	37.40	1	37.42	1
GLU	161	HG2/HG3	2.27	2	2.25	2
GLU	161	C	175.97	1		
VAL	162	N	108.38	1	108.40	1
VAL	162	HN	6.85	1	6.85	1
VAL	162	CA	57.07	1	56.98	1
VAL	162	HA	4.86	1	4.86	1
VAL	162	CB	34.81	1	34.81	1
VAL	162	HB	2.19	1	2.20	1
VAL	162	CG1/CG2	22.74/19.76	2	22.67/19.80	2
VAL	162	HG1/HG2	1.06/0.76	2	1.06/0.78	2
PRO	163	CA	65.21	1	65.06	1
PRO	163	HA	4.15	1	4.14	1
PRO	163	CB	32.41	1	32.44	1
PRO	163	HB2/HB3	1.70/2.14	2	1.70/2.13	2
PRO	163	CG	27.73	1	27.60	1
PRO	163	HG2/HG3	2.00/1.88	2	1.99/1.88	2
PRO	163	CD	51.13	1	50.99	1
PRO	163	HD2/HD3	4.02/3.59	2	4.02/3.61	2
PRO	163	C	178.72	1		
GLU	164	N	122.28	1	122.22	1
GLU	164	HN	9.45	1	9.48	1
GLU	164	CA	62.07	1	61.90	1
GLU	164	HA	3.31	1	3.30	1
GLU	164	CB	28.25	1	28.25	1
GLU	164	HB2/HB3	0.74	2	0.77	2
GLU	164	HG2/HG3	2.63/2.37	2		
GLU	164	C	176.87	1		

Residue name	Residue number	Atom name	Chemical shift [ppm]	Ambiguity	Chemical shift [ppm]	Ambiguity
			FIR RRM1-RRM2 free		FIR RRM1-RRM2 bound	
ALA	165	N	118.62	1	118.53	1
ALA	165	HN	6.33	1	6.31	1
ALA	165	CA	54.35	1	54.40	1
ALA	165	HA	3.65	1	3.63	1
ALA	165	CB	19.97	1	19.92	1
ALA	165	HB	0.66	1	0.65	1
ALA	165	C	177.15	1		
ALA	166	N	117.01	1	116.95	1
ALA	166	HN	5.94	1	5.92	1
ALA	166	CA	54.38	1	54.41	1
ALA	166	HA	3.66	1	3.64	1
ALA	166	CB	18.62	1	18.62	1
ALA	166	HB	1.47	1	1.46	1
ALA	166	C	178.14			
GLN	167	N	115.49	1	115.53	1
GLN	167	HN	7.38	1	7.41	1
GLN	167	CA	58.32	1	58.23	1
GLN	167	HA	3.72	1	3.69	1
GLN	167	CB	27.46	1		
GLN	167	HB2/HB3	1.98	2	1.95	2
GLN	167	CG	32.61	1	32.43	1
GLN	167	HG2/HG3	2.18/2.35	2	2.16/2.33	2
GLN	167	NE2	113.15	1	113.20	1
GLN	167	HE21/HE22	6.71/8.04	2	6.71/8.02	2
GLN	167	C	177.48	1		
LEU	168	N	117.01	1	116.87	1
LEU	168	HN	7.47	1	7.46	1
LEU	168	CA	57.68	1	57.61	1
LEU	168	HA	4.03	1	4.02	1
LEU	168	CB	42.89	1	42.76	1
LEU	168	HB2/HB3	1.55/1.87	2	1.53/1.87	2
LEU	168	CG	26.65	1	26.78	1
LEU	168	HG	1.89	1	1.90	1
LEU	168	CD1/CD2	24.13/26.52	2	24.00/26.41	2
LEU	168	HD1/HD2	0.91/1.03	2	0.90/1.01	2
LEU	168	C	178.23	1		
ALA	169	N	121.42	1	121.45	1
ALA	169	HN	7.73	1	7.71	1
ALA	169	CA	55.19	1	55.09	1
ALA	169	HA	2.38	1	2.39	1
ALA	169	CB	18.31	1	18.22	1
ALA	169	HB	1.35	1	1.34	1
ALA	169	C	179.51	1		
LEU	170	N	117.01	1	117.04	1
LEU	170	HN	8.00	1	8.00	1
LEU	170	CA	58.37	1	58.30	1
LEU	170	HA	3.74	1	3.74	1
LEU	170	CB	42.10	1	42.07	1
LEU	170	HB2/HB3	1.57/1.79	2	1.58/1.77	2
LEU	170	CG	26.92	1	26.88	1
LEU	170	HG	1.72	1	1.71	1
LEU	170	CD1/CD2	24.89/25.30	2	24.92/25.38	2
LEU	170	HD1/HD2	0.81/0.80	2	0.80/0.79	2
LEU	170	C	179.21	1		
GLU	171	N	116.11	1	116.04	1
GLU	171	HN	7.60	1	7.59	1
GLU	171	CA	58.60	1	58.65	1
GLU	171	HA	4.04	1	4.03	1
GLU	171	CB	30.76	1	30.70	1
GLU	171	HB2/HB3	2.19/2.06	2	2.17/2.05	2
GLU	171	CG	36.83	1	36.83	1
GLU	171	HG2/HG3	2.27/2.66	2	2.23/2.65	2
GLU	171	C	179.25	1		
GLN	172	N	114.55	1	114.52	1
GLN	172	HN	8.16	1	8.18	1
GLN	172	CA	57.72	1	57.46	1
GLN	172	HA	4.46	1	4.45	1
GLN	172	CB	29.61	1	29.50	1
GLN	172	HB2/HB3	2.37/1.75	2	2.38/1.75	2
GLN	172	CG	34.83	1	34.78	1
GLN	172	HG2/HG3	2.41/2.77	2	2.39/2.77	2
GLN	172	NE2	110.05	1	110.25	1
GLN	172	HE21/HE22	7.75/6.90	2	7.76/6.86	2
GLN	172	C	178.02	1		
MET	173	N	115.39	1	115.23	1
MET	173	HN	8.09	1	8.10	1
MET	173	CA	54.02	1	53.93	1
MET	173	HA	5.04	1	5.04	1

Residue name	Residue number	Atom name	Chemical shift [ppm]	Ambiguity	Chemical shift [ppm]	Ambiguity
			FIR RRM1-RRM2 free		FIR RRM1-RRM2 bound	
MET	173	CB	31.12	1	31.02	1
MET	173	HB2/HB3	1.94/2.14	2	1.91/2.13	2
MET	173	CE	16.42	1	16.31	1
MET	173	HE	1.39	1	1.38	1
MET	173	C	176.54	1		
ASN	174	N	116.23	1	116.13	1
ASN	174	HN	7.22	1	7.19	1
ASN	174	CA	55.93	1	55.83	1
ASN	174	HA	4.48	1	4.46	1
ASN	174	CB	39.01	1	38.98	1
ASN	174	HB2/HB3	2.88/2.97	2	2.88/2.96	2
SER	175	CA	59.98	1	59.96	1
SER	175	HA			4.34	1
SER	175	CB	63.13	1	62.94	1
SER	175	HB2/HB3			4.02/4.05	2
SER	175	C	174.05	1		
VAL	176	N	120.71	1	120.60	1
VAL	176	HN	7.40	1	7.38	1
VAL	176	CA	62.66	1	62.56	1
VAL	176	HA	4.06	1	4.04	1
VAL	176	CB	32.86	1	32.80	1
VAL	176	HB	2.08	1	2.07	1
VAL	176	CG1/CG2	22.01/21.06	2	21.94/21.05	2
VAL	176	HG1/HG2	1.03/0.90	2	1.02/0.89	2
VAL	176	C	174.84	1		
MET	177	N	123.11	1	123.10	1
MET	177	HN	8.27	1	8.26	1
MET	177	CA	55.20	1	54.95	1
MET	177	HA	4.62	1	4.59	1
MET	177	CB	33.36	1	33.29	1
MET	177	HB2/HB3	1.92/1.79	2	1.91/1.79	2
MET	177	CG	32.26	1	32.29	1
MET	177	HG2/HG3	2.32/2.28	2	2.29/2.28	2
MET	177	CE	16.98	1	16.97	1
MET	177	HE	1.93	1	1.91	1
MET	177	C	175.49	1		
LEU	178	N	126.36	1	126.52	1
LEU	178	HN	8.54	1	8.54	1
LEU	178	CA	54.40	1	54.16	1
LEU	178	HA	4.65	1	4.63	1
LEU	178	CB	44.74	1	44.67	1
LEU	178	HB2/HB3	1.58/1.42	2	1.57/1.41	2
LEU	178	CD1/CD2	24.60	2	24.71	2
LEU	178	HD1/HD2	0.82	2	0.82	2
GLY	180	CA	45.21	1	45.09	1
GLY	180	C	173.59	1		
ARG	181	N	118.87	1	118.84	1
ARG	181	HN	7.41	1	7.39	1
ARG	181	CA	54.43	1	54.29	1
ARG	181	CB	33.38	1		
ARG	181	HB2/HB3	1.90/1.77	2	1.87/1.78	2
ARG	181	CG	27.05	1	26.93	1
ARG	181	HG2/HG3	1.54	2	1.54	2
ARG	181	CD	43.37	1	43.21	1
ARG	181	HD2/HD3	3.01	2	3.01	2
ARG	181	C	174.24	1		
ASN	182	N	120.54	1	120.43	1
ASN	182	HN	8.34	1	8.34	1
ASN	182	CA	54.24	1	54.08	1
ASN	182	HA	4.86	1	4.85	1
ASN	182	CB	39.72	1	39.56	1
ASN	182	HB2/HB3	2.52	2	2.50	2
ASN	182	C	175.28	1		
ILE	183	N	118.06	1	118.09	1
ILE	183	HN	8.34	1	8.34	1
ILE	183	CA	60.99	1	60.77	1
ILE	183	HA	4.65	1	4.65	1
ILE	183	CB	40.34	1	40.22	1
ILE	183	HB	2.02	1	2.02	1
ILE	183	CG1	25.22	1	25.15	1
ILE	183	HG12/HG13	1.50	2	1.50	2
ILE	183	CG2	19.17	1	19.19	1
ILE	183	HG2	0.95	1	0.94	1
ILE	183	CD1	14.80	1	14.83	1
ILE	183	HD1	0.87	1	0.87	1
ILE	183	C	175.31	1		
LYS	184	N	121.34	1	121.35	1
LYS	184	HN	7.97	1	7.96	1

Residue name	Residue number	Atom name	Chemical shift [ppm]	Ambiguity	Chemical shift [ppm]	Ambiguity
			FIR RRM1-RRM2 free		FIR RRM1-RRM2 bound	
LYS	184	CA	54.25	1	54.21	1
LYS	184	HA	4.95	1	4.95	1
LYS	184	CB	35.15	1	35.10	1
LYS	184	HB2/HB3	1.60/1.74	2	1.59/1.74	2
LYS	184	CG	24.75	1	24.69	1
LYS	184	HG2/HG3	1.45/1.36	2	1.44/1.36	2
LYS	184	CD	29.24	1	29.03	1
LYS	184	HD2/HD3	1.73	2	1.73	2
LYS	184	CE	41.70	1	41.88	1
LYS	184	HE2/HE3	3.02	2	2.99	2
LYS	184	C	175.30	1		
VAL	185	N	121.79	1	121.77	1
VAL	185	HN	8.60	1	8.60	1
VAL	185	CA	60.25	1	60.09	1
VAL	185	HA	5.36	1	5.36	1
VAL	185	CB	34.58	1	34.46	1
VAL	185	HB	1.90	1	1.88	1
VAL	185	CG1/CG2	22.42	2	22.49	2
VAL	185	HG1/HG2	1.00	2	0.99	2
VAL	185	C	175.53	1		
GLY	186	N	112.45	1	112.53	1
GLY	186	HN	9.00	1	8.99	1
GLY	186	CA	44.99	1	44.81	1
GLY	186	HA2/HA3	4.50/3.93	2	4.48/3.89	2
ARG	187	CA			53.75	1
ARG	187	HA			4.84	1
SER	189	CA			58.25	1
SER	189	HA			4.45	1
SER	189	CB			62.91	1
SER	189	HB2/HB3			4.74	2
ASN	190	CA	52.87	1	52.74	1
ASN	190	HA	4.77	1	4.75	1
ASN	190	CB	36.70	1	36.54	1
ASN	190	HB2/HB3	2.74/2.99	2	2.73/2.98	2
ILE	191	HN			7.59	1
ILE	191	CA	62.35	1	62.26	1
ILE	191	HA	4.02	1	4.02	1
ILE	191	CB	38.92	1	38.63	1
ILE	191	HB	1.88	1	1.88	1
ILE	191	CG1	28.55	1	28.51	1
ILE	191	HG12/HG13	1.24/1.43	2	1.24/1.43	2
ILE	191	CG2	18.12	1	18.09	1
ILE	191	HG2	0.89	1	0.88	1
ILE	191	CD1	14.88	1	14.89	1
ILE	191	HD1	0.85	1	0.85	1
GLY	192	N	109.86	1	109.84	1
GLY	192	HN	8.52	1	8.52	1
GLY	192	CA	47.13	1	47.05	1
GLN	193	CA	57.59	1	57.42	1
GLN	193	HA	4.19	1	4.17	1
GLN	193	CB	28.42	1	28.41	1
GLN	193	HB2/HB3	2.17/2.09	2	2.15/2.06	2
GLN	193	CG	34.09	1	33.93	1
GLN	193	HG2/HG3	2.43	2	2.41	2
ALA	194	N	118.98	1	118.99	1
ALA	194	HN	7.71	1	7.66	1
ALA	194	CA	52.02	1	51.80	1
ALA	194	HA	4.51	1	4.49	1
ALA	194	CB	19.80	1	19.69	1
ALA	194	HB	1.44	1	1.43	1
GLN	195	N	120.89	1	120.88	1
GLN	195	HN	7.77	1	7.76	1
GLN	195	CA	60.19	1	60.17	1
GLN	195	CG	33.35	1	33.36	1
GLN	195	HG2/HG3	2.38	2	2.37	2
GLN	195	NE2	114.96	1	114.98	1
GLN	195	HE21/HE22	6.56/7.91	2	6.57/7.90	2
PRO	196	CA	66.45	1	66.25	1
PRO	196	HA	4.44	1	4.43	1
PRO	196	CB	31.16	1	31.09	1
PRO	196	HB2/HB3	1.76/2.39	2	1.75/2.38	2
PRO	196	CG	28.55	1	28.57	1
PRO	196	HG2/HG3	2.08/2.00	2	2.06/1.99	2
PRO	196	CD	50.02	1	50.02	1
PRO	196	HD2/HD3	3.51/3.69	2	3.50/3.68	2
PRO	196	C	179.60	1		
ILE	197	N	116.68	1	116.75	1
ILE	197	HN	7.15	1	7.14	1

Residue name	Residue number	Atom name	Chemical shift [ppm]	Ambiguity	Chemical shift [ppm]	Ambiguity
			FIR RRM1-RRM2 free		FIR RRM1-RRM2 bound	
ILE	197	CA	62.89	1	62.79	1
ILE	197	HA	3.89	1	3.88	1
ILE	197	CB	37.09	1	37.05	1
ILE	197	HB	2.00	1	1.99	1
ILE	197	CG1	28.54	1	28.54	1
ILE	197	HG12/HG13	1.33/1.47	2	1.32/1.46	2
ILE	197	CG2	18.49	1	18.45	1
ILE	197	HG2	0.88	1	0.87	1
ILE	197	CD1	11.80	1	11.75	1
ILE	197	HD1	0.79	1	0.80	1
ILE	197	C	177.55	1		
ILE	198	N	123.20	1	123.13	1
ILE	198	HN	7.97	1	7.97	1
ILE	198	CA	66.62	1	66.39	1
ILE	198	HA	3.55	1	3.55	1
ILE	198	CB	37.66	1	37.69	1
ILE	198	HB	2.00	1	1.99	1
ILE	198	CG1	28.38	1	28.30	1
ILE	198	HG12/HG13	1.22/1.63	2	1.23/1.61	2
ILE	198	CG2	16.37	1	16.31	1
ILE	198	HG2	0.93	1	0.92	1
ILE	198	CD1	13.97	1	13.88	1
ILE	198	HD11	0.74	1	0.73	1
ILE	198	C	179.21	1		
ASP	199	N	119.81	1	119.82	1
ASP	199	HN	8.65	1	8.64	1
ASP	199	CA	57.43	1	57.29	1
ASP	199	HA	4.40	1	4.38	1
ASP	199	CB	39.90	1	39.82	1
ASP	199	HB2/HB3	2.68/2.70	2	2.65/2.69	2
ASP	199	C	179.12	1		
GLN	200	N	122.15	1	122.15	1
GLN	200	HN	7.65	1	7.65	1
GLN	200	CA	59.24	1	59.14	1
GLN	200	HA	4.14	1	4.12	1
GLN	200	CB	28.48	1	28.51	1
GLN	200	HB2/HB3	2.26/2.20	2	2.25/2.19	2
GLN	200	CG	33.52	1	33.39	1
GLN	200	HG2/HG3	2.56/2.35	2	2.55/2.34	2
GLN	200	NE2	111.17	1	111.26	1
GLN	200	HE21/HE22	6.68/7.65	2	6.69/7.65	2
GLN	200	C	178.58	1		
LEU	201	N	120.44	1	120.37	1
LEU	201	HN	8.64	1	8.63	1
LEU	201	CA	58.44	1	58.28	1
LEU	201	HA	3.98	1	3.96	1
LEU	201	CB	41.91	1	41.80	1
LEU	201	HB2/HB3	1.43/2.15	2	1.45/2.14	2
LEU	201	CG	27.32	1	27.39	1
LEU	201	HG	1.86	1	1.86	1
LEU	201	CD1/CD2	26.71/24.08	2	26.71/24.00	2
LEU	201	HD1/HD2	0.80/0.81	2	0.79/0.81	2
LEU	201	C	178.56	1		
ALA	202	N	120.94	1	120.86	1
ALA	202	HN	8.00	1	7.98	1
ALA	202	CA	55.12	1	55.00	1
ALA	202	HA	3.94	1	3.92	1
ALA	202	CB	17.88	1	17.82	1
ALA	202	HB	1.57	1	1.56	1
ALA	202	C	180.80			
GLU	203	N	119.76	1	119.72	1
GLU	203	HN	7.81	1	7.80	1
GLU	203	CA	59.20	1	59.16	1
GLU	203	HA	4.07	1	4.06	1
GLU	203	CB	29.00	1	29.17	1
GLU	203	HB2/HB3	2.18	2	2.18	2
GLU	203	CG	36.00	1	35.81	1
GLU	203	HG2/HG3	2.32/2.30	2	2.31/2.30	2
GLU	203	C	180.10	1		
GLU	204	N	120.61	1	120.57	1
GLU	204	HN	8.48	1	8.47	1
GLU	204	CA	59.20	1	59.21	1
GLU	204	HA	4.06	1	4.05	1
GLU	204	CB	30.34	1	30.18	1
GLU	204	HB2/HB3	2.07/2.15	2	2.07/2.15	2
GLU	204	CG	36.76	1	36.65	1
GLU	204	HG2/HG3	2.30/2.67	2	2.30/2.66	2
GLU	204	C	179.76	1		

Residue name	Residue number	Atom name	Chemical shift [ppm]	Ambiguity	Chemical shift [ppm]	Ambiguity
			FIR RRM1-RRM2 free		FIR RRM1-RRM2 bound	
ALA	205	N	119.03	1	119.07	1
ALA	205	HN	8.23	1	8.20	1
ALA	205	CA	54.64	1	54.59	1
ALA	205	HA	3.87	1	3.87	1
ALA	205	CB	18.54	1	18.49	1
ALA	205	HB	1.15	1	1.14	1
ALA	205	C	179.50	1		
ARG	206	N	116.58	1	116.51	1
ARG	206	HN	7.48	1	7.47	1
ARG	206	CA	59.17	1	59.05	1
ARG	206	HA	4.20	1	4.19	1
ARG	206	CB	30.47	1	30.38	1
ARG	206	HB2/HB3	1.99	2	1.99	2
ARG	206	CG	27.97	1	27.93	1
ARG	206	HG2/HG3	2.01/1.76	2	2.00/1.75	2
ARG	206	CD	43.87	1	43.67	1
ARG	206	HD2/HD3	3.28	2	3.26	2
ARG	206	C	177.20	1		
ALA	207	N	118.65	1	118.62	1
ALA	207	HN	7.31	1	7.30	1
ALA	207	CA	53.29	1	53.17	1
ALA	207	HA	4.12	1	4.12	1
ALA	207	CB	18.64	1	18.65	1
ALA	207	HB	1.16	1	1.15	1
ALA	207	C	176.56	1		
PHE	208	N	114.60	1	114.62	1
PHE	208	HN	7.71	1	7.70	1
PHE	208	CA	57.46	1	57.33	1
PHE	208	HA	4.48	1	4.49	1
PHE	208	CB	41.43	1		
PHE	208	HB2/HB3	3.02/2.93	2	3.00/2.93	2
PHE	208	CD1/CD2	132.25	3	132.38	3
PHE	208	HD1/HD2	7.39	3	7.40	3
PHE	208	CE1/CE2	131.81	3	131.75	3
PHE	208	HE1/HE2	7.37	3	7.37	3
ASN	209	CA	52.65	1	52.39	1
ASN	209	HA	4.91	1	4.88	1
ASN	209	CB	37.36	1	37.15	1
ASN	209	HB2/HB3	2.65/3.57	2	2.67/3.57	2
ASN	209	ND2			113.48	1
ASN	209	HD21/HD22	6.62	2	6.62	2
ASN	209	C	174.55	1		
ARG	210	N	119.71	1	119.62	1
ARG	210	HN	8.24	1	8.28	1
ARG	210	CA	55.09	1	55.07	1
ARG	210	HA	5.74	1	5.73	1
ARG	210	CB	34.55	1	34.34	1
ARG	210	HB2/HB3	1.63/1.87	2	1.65/1.85	2
ARG	210	CG	29.12	1	29.05	1
ARG	210	HG2/HG3	1.48/1.73	2	1.47/1.71	2
ARG	210	HD2/HD3	3.18	2	3.16	2
ARG	210	C	176.11	1		
ILE	211	N	118.56	1	118.69	1
ILE	211	HN	9.51	1	9.53	1
ILE	211	CA	59.36	1	59.20	1
ILE	211	HA	5.08	1	5.08	1
ILE	211	CB	40.99	1	40.94	1
ILE	211	HB	1.99	1	1.99	1
ILE	211	CG1	26.74	1	26.74	1
ILE	211	HG12/HG13	1.07/1.26	2	1.03/1.24	2
ILE	211	CG2	19.24	1	19.15	1
ILE	211	HG2	0.97	1	0.96	1
ILE	211	CD1	15.34	1	15.16	1
ILE	211	HD1	0.75	1	0.74	1
ILE	211	C	172.15	1		
TYR	212	N	124.92	1	124.94	1
TYR	212	HN	9.16	1	9.16	1
TYR	212	CA	55.65	1	55.65	1
TYR	212	HA	4.72	1	4.72	1
TYR	212	CB	42.75	1	42.65	1
TYR	212	HB2/HB3	2.34/2.53	2	2.32/2.50	2
TYR	212	CD1/CD2	133.19/132.65	3	133.02/132.65	3
TYR	212	HD1/HD2	6.46/5.99	3	6.46/5.95	3
TYR	212	CE1/CE2	116.56/118.16	3	116.55/118.11	3
TYR	212	HE1/HE2	6.56/6.30	3	6.56/6.32	3
TYR	212	HH	9.08	1	9.09	1
TYR	212	C	172.36	1		
VAL	213	N	124.11	1	123.77	1

Residue name	Residue number	Atom name	Chemical shift [ppm]	Ambiguity	Chemical shift [ppm]	Ambiguity
			FIR RRM1-RRM2 free		FIR RRM1-RRM2 bound	
VAL	213	HN	7.69	1	7.64	1
VAL	213	CA	59.03	1	58.94	1
VAL	213	HA	4.57	1	4.59	1
VAL	213	CB	34.57	1	34.60	1
VAL	213	HB	1.25	1	1.22	1
VAL	213	CG1/CG2	21.37/20.82	2	21.12/20.85	2
VAL	213	HG1/HG2	0.38/0.44	2	0.35/0.40	2
VAL	213	C	171.69	1		
ALA	214	N	126.77	1	126.60	1
ALA	214	HN	8.98	1	8.93	1
ALA	214	CA	50.20	1	50.15	1
ALA	214	HA	5.10	1	5.10	1
ALA	214	CB	22.50	1	22.59	1
ALA	214	HB	1.27	1	1.27	1
ALA	214	C	176.84	1		
SER	215	N	111.09	1	111.09	1
SER	215	HN	8.29	1	8.25	1
SER	215	CA	60.80	1	60.70	1
SER	215	HA	3.85	1	3.85	1
SER	215	CB	63.48	1	63.06	1
SER	215	HB2/HB3	4.44/4.27	2	4.45/4.19	2
SER	215	C	174.43	1		
VAL	216	N	118.70	1	118.51	1
VAL	216	HN	8.07	1	8.11	1
VAL	216	CA	62.34	1	62.42	1
VAL	216	HA	3.54	1	3.45	1
VAL	216	CB	31.73	1	31.87	1
VAL	216	HB	1.29	1	1.28	1
VAL	216	CG1/CG2	21.76/21.40	2	21.69/21.42	2
VAL	216	HG1/HG2	0.67/0.38	2	0.67/0.37	2
VAL	216	C	176.50	1		
HIS	217	N	131.87	1	131.60	1
HIS	217	HN	9.10	1	8.92	1
HIS	217	CA	59.94	1	59.81	1
HIS	217	HA	3.90	1	3.89	1
HIS	217	CB	31.61	1	31.59	1
HIS	217	HB2/HB3	2.84/3.09	2	2.79/3.07	2
HIS	217	CD2	118.36	1	118.10	1
HIS	217	HD2	6.96	1	7.04	1
HIS	217	CE1	138.97	1	138.82	1
HIS	217	HE1	7.71	1	7.78	1
HIS	217	C	176.62	1		
GLN	218	N	125.42	1	126.03	1
GLN	218	HN	7.93	1	8.07	1
GLN	218	CA	58.47	1	58.34	1
GLN	218	HA	3.63	1	3.63	1
GLN	218	CB	29.33	1	29.30	1
GLN	218	HB2/HB3	1.83/1.78	2	1.85/1.79	2
GLN	218	CG	33.47	1	33.39	1
GLN	218	HG2/HG3	2.19/2.07	2	2.21/2.11	2
GLN	218	NE2	113.01	1	113.03	1
GLN	218	HE21/HE22	7.40/6.71	2	7.40/6.72	2
GLN	218	C	176.20	1		
ASP	219	N	121.44	1	121.76	1
ASP	219	HN	10.83	1	11.16	1
ASP	219	CA	55.69	1	55.49	1
ASP	219	HA	4.73	1	4.75	1
ASP	219	CB	40.59	1	40.88	1
ASP	219	HB2/HB3	2.72/2.87	2	2.72/2.88	2
ASP	219	C	177.10	1		
LEU	220	N	121.23	1	120.37	1
LEU	220	HN	7.92	1	7.87	1
LEU	220	CA	54.47	1	54.06	1
LEU	220	HA	4.63	1	4.60	1
LEU	220	CB	43.04	1	43.01	1
LEU	220	HB2/HB3	2.00/1.28	2	1.99/1.28	2
LEU	220	CG	27.36	1	27.49	1
LEU	220	HG	1.67	1	1.60	1
LEU	220	CD1/CD2	22.86/26.98	2	23.43/27.19	2
LEU	220	HD1/HD2	0.80/0.93	2	0.69/0.91	2
LEU	220	C	176.09	1		
SER	221	N	116.03	1	115.92	1
SER	221	HN	9.47	1	9.36	1
SER	221	CA	56.07	1	55.84	1
SER	221	HA	4.77	1	4.77	1
SER	221	CB	67.11	1	67.04	1
SER	221	HB2/HB3	3.99/4.21	2	3.95/4.20	2
SER	221	C	175.39	1		

Residue name	Residue number	Atom name	Chemical shift [ppm]	Ambiguity	Chemical shift [ppm]	Ambiguity
			FIR RRM1-RRM2 free		FIR RRM1-RRM2 bound	
ASP	222	N	118.44	1	118.22	1
ASP	222	HN	9.19	1	9.26	1
ASP	222	CA	57.74	1	57.47	1
ASP	222	HA	3.74	1	3.70	1
ASP	222	CB	39.62	1	39.21	1
ASP	222	HB2/HB3	2.61	2	2.63/2.59	2
ASP	222	C	178.08	1		
ASP	223	N	116.07	1	115.97	1
ASP	223	HN	8.31	1	8.27	1
ASP	223	CA	57.20	1	57.14	1
ASP	223	HA	4.40	1	4.36	1
ASP	223	CB	40.44	1	40.56	1
ASP	223	HB2/HB3	2.65/2.51	2	2.60/2.50	2
ASP	223	C	178.67	1		
ASP	224	N	120.87	1	121.86	1
ASP	224	HN	7.62	1	7.63	1
ASP	224	CA	57.51	1	57.45	1
ASP	224	HA	4.41	1	4.36	1
ASP	224	CB	41.98	1	41.76	1
ASP	224	HB2/HB3	2.99/2.68	2	3.00/2.51	2
ASP	224	C	178.68	1		
ILE	225	N	117.30	1	117.02	1
ILE	225	HN	7.35	1	7.33	1
ILE	225	CA	61.50	1	61.95	1
ILE	225	HA	3.95	1	3.83	1
ILE	225	CB	35.92	1	35.71	1
ILE	225	HB	1.89	1	1.83	1
ILE	225	CG1	27.10	1	27.19	1
ILE	225	HG12/HG13	1.06/1.51	2	1.03/1.52	2
ILE	225	CG2	18.34	1	18.21	1
ILE	225	HG2	0.64	1	0.62	1
ILE	225	CD1	10.76	1	10.67	1
ILE	225	HD1	0.63	1	0.63	1
ILE	225	C	177.57	1		
LYS	226	N	123.48	1	123.11	1
LYS	226	HN	8.47	1	8.37	1
LYS	226	CA	61.06	1	60.86	1
LYS	226	HA	3.70	1	3.67	1
LYS	226	CB	32.79	1	32.85	1
LYS	226	HB2/HB3	1.99/2.02	2	1.98	2
LYS	226	CG	25.52	1	25.29	1
LYS	226	HG2/HG3	1.26/1.39	2	1.19/1.34	2
LYS	226	CD	29.72	1	29.55	1
LYS	226	HD2/HD3	1.68	2	1.64	2
LYS	226	CE	42.08	1	41.68	1
LYS	226	HE2/HE3	2.86	2	2.84	2
LYS	226	C	177.72	1		
SER	227	N	110.69	1	111.71	1
SER	227	HN	7.82	1	7.92	1
SER	227	CA	61.77	1	61.90	1
SER	227	HA	4.24	1	4.21	1
SER	227	CB	63.25	1	63.29	1
SER	227	HB2/HB3	4.01	2	4.11/4.02	2
SER	227	C	176.78	1		
VAL	228	N	117.13	1	118.16	1
VAL	228	HN	7.19	1	7.27	1
VAL	228	CA	64.82	1	66.36	1
VAL	228	HA	3.93	1	3.57	1
VAL	228	CB	32.28	1	32.40	1
VAL	228	HB	2.09	1	2.07	1
VAL	228	CG1/CG2	21.90/21.72	2	22.81/21.68	2
VAL	228	HG1/HG2	1.05/0.67	2	0.98/0.47	2
VAL	228	C	177.93	1		
PHE	229	N	117.02	1	114.78	1
PHE	229	HN	8.27	1	8.32	1
PHE	229	CA	62.13	1	62.54	1
PHE	229	HA	4.41	1	4.18	1
PHE	229	CB	37.91	1	37.53	1
PHE	229	HB2/HB3	3.53/2.83	2	3.47/2.81	2
PHE	229	CD1/CD2	132.46	3	132.08	3
PHE	229	HD1/HD2	7.78	3	7.68	3
PHE	229	CE1/CE2	130.18	3	129.94	3
PHE	229	HE1/HE2	6.82	3	6.76	3
PHE	229	CZ	128.72	1	128.90	1
PHE	229	HZ	6.90	1	6.87	1
PHE	229	C	179.49	1		
GLU	230	N	126.27	1	126.36	1
GLU	230	HN	8.76	1	8.91	1

Residue name	Residue number	Atom name	Chemical shift [ppm]	Ambiguity	Chemical shift [ppm]	Ambiguity
			FIR RRM1-RRM2 free		FIR RRM1-RRM2 bound	
GLU	230	CA	58.22	1	57.93	1
GLU	230	HA	4.87	1	5.05	1
GLU	230	CB	28.89	1	28.73	1
GLU	230	HB2/HB3	2.31/2.11	2	2.29/2.08	2
GLU	230	CG	37.62	1	37.98	1
GLU	230	HG2/HG3	2.54	2	2.56	2
GLU	230	C	177.08	1		
ALA	231	N	121.44	1	122.64	1
ALA	231	HN	7.08	1	7.46	1
ALA	231	CA	54.23	1	54.60	1
ALA	231	HA	3.91	1	3.87	1
ALA	231	CB	17.99	1	18.37	1
ALA	231	HB	0.80	1	0.93	1
ALA	231	C	178.45	1		
PHE	232	N	111.24	1	109.82	1
PHE	232	HN	6.88	1	7.09	1
PHE	232	CA	58.50	1	58.74	1
PHE	232	HA	4.39	1	4.25	1
PHE	232	CB	39.28	1	39.36	1
PHE	232	HB2/HB3	3.48/2.65	2	3.44/2.72	2
PHE	232	CD1/CD2	131.56	3	131.54	3
PHE	232	HD1/HD2	7.24	3	7.35	3
PHE	232	CE1/CE2	131.41	3	131.39	3
PHE	232	HE1/HE2	7.37	3	7.05	3
PHE	232	C	175.32	1		
GLY	233	N	108.64	1	108.81	1
GLY	233	HN	7.62	1	7.61	1
GLY	233	CA	44.25	1	44.29	1
GLY	233	HA2/HA3	3.89/4.50	2	3.89/4.53	2
GLY	233	C	170.49	1		
LYS	234	N	116.61	1	116.52	1
LYS	234	HN	8.05	1	8.04	1
LYS	234	CA	57.65	1	57.82	1
LYS	234	HA	4.30	1	4.26	1
LYS	234	CB	32.93	1	32.99	1
LYS	234	HB2/HB3	1.78	2	1.78	2
LYS	234	CG	24.96	1	24.99	1
LYS	234	HG2/HG3	1.48	2	1.48/1.38	2
LYS	234	CD	29.14	1	28.91	1
LYS	234	HD2/HD3	1.72/1.60	2	1.72/1.59	2
LYS	234	C	177.59	1		
ILE	235	N	128.28	1	128.25	1
ILE	235	HN	9.11	1	9.28	1
ILE	235	CA	61.96	1	61.90	1
ILE	235	HA	4.03	1	4.06	1
ILE	235	CB	40.36	1	40.35	1
ILE	235	HB	1.64	1	1.61	1
ILE	235	CG1	29.98	1	30.12	1
ILE	235	HG12/HG13	0.18/1.96	2	0.03/1.93	2
ILE	235	CG2	18.05	1	18.14	1
ILE	235	HG2	0.58	1	0.58	1
ILE	235	CD1	14.66	1	14.78	1
ILE	235	HD1	0.70	1	0.68	1
ILE	235	C	176.00	1		
LYS	236	N	128.21	1	128.04	1
LYS	236	HN	9.10	1	9.00	1
LYS	236	CA	57.76	1	57.84	1
LYS	236	HA	4.31	1	4.29	1
LYS	236	CB	33.80	1	33.84	1
LYS	236	HB2/HB3	1.72/1.47	2	1.69/1.49	2
LYS	236	CG	24.92	1	24.90	1
LYS	236	HG2/HG3	1.36	2	1.35	2
LYS	236	HE2/HE3	2.97	2	2.96	2
LYS	236	C	176.44	1		
SER	237	N	110.95	1	110.95	1
SER	237	HN	7.45	1	7.44	1
SER	237	CA	57.70	1	57.57	1
SER	237	HA	4.48	1	4.50	1
SER	237	CB	65.08	1		
CYS	238	CA	58.38	1	58.09	1
CYS	238	HA	4.76	1	4.77	1
CYS	238	CB	26.72	1	26.52	1
CYS	238	HB2/HB3	2.35/1.84	2	2.32/2.80	2
CYS	238	C	172.14	1		
THR	239	N	121.56	1	121.81	1
THR	239	HN	8.16	1	8.18	1
THR	239	CA	60.44	1	60.23	1
THR	239	HA	4.86	1	4.87	1

Residue name	Residue number	Atom name	Chemical shift [ppm]	Ambiguity	Chemical shift [ppm]	Ambiguity
			FIR RRM1-RRM2 free		FIR RRM1-RRM2 bound	
THR	239	CB	72.12	1	71.79	1
THR	239	HB	3.95	1	3.93	1
THR	239	CG2	21.80	1	21.75	1
THR	239	HG2	1.18	1	1.17	1
THR	239	C	173.89	1		
LEU	240	N	126.89	1	126.87	1
LEU	240	HN	8.62	1	8.61	1
LEU	240	CA	53.57	1	53.26	1
LEU	240	HA	4.57	1	4.58	1
LEU	240	CB	41.56	1	41.45	1
LEU	240	HB2/HB3	0.80/1.49	2	0.76/1.47	2
LEU	240	CG	27.01	1	26.94	1
LEU	240	HG	1.16	1	1.13	1
LEU	240	CD1/CD2	24.68/21.59	2	24.61/21.47	2
LEU	240	HD1/HD2	-0.05/0.11	2	-0.09/0.08	2
LEU	240	C	176.73	1		
ALA	241	N	127.79	1	127.75	1
ALA	241	HN	8.45	1	8.36	1
ALA	241	CA	52.47	1	52.42	1
ALA	241	HA	4.14	1	4.13	1
ALA	241	CB	19.18	1	19.30	1
ALA	241	HB	0.78	1	0.78	1
ALA	241	C	177.27	1		
ARG	242	N	120.89	1	120.88	1
ARG	242	HN	8.46	1	8.47	1
ARG	242	CA	54.87	1	54.71	1
ARG	242	HA	4.73	1	4.71	1
ARG	242	CB	33.34	1	33.22	1
ARG	242	HB2/HB3	1.62/1.37	2	1.62/1.36	2
ARG	242	CG	27.26	1	27.20	1
ARG	242	HG2/HG3	1.59/1.34	2	1.58/1.33	2
ARG	242	CD	42.94	1	42.81	1
ARG	242	HD2/HD3	2.96/3.04	2	2.96/3.03	2
ARG	242	C	175.48	1		
ASP	243	N	123.72	1	123.65	1
ASP	243	HN	8.44	1	8.43	1
ASP	243	CA	51.01	1	50.90	1
ASP	243	HA	4.92	1	4.91	1
ASP	243	CB	42.76	1	42.68	1
ASP	243	HB2/HB3	3.16/2.41	2	3.16/2.40	2
PRO	244	CA	64.36	1	64.26	1
PRO	244	HA	4.39	1	4.38	1
PRO	244	CB	32.34	1	32.27	1
PRO	244	HB2/HB3	2.36/2.03	2	2.34/2.03	2
PRO	244	CG	27.20	1	27.11	1
PRO	244	HG2/HG3	2.07	2	2.06	2
PRO	244	CD	51.25	1	51.18	1
PRO	244	HD2/HD3	4.00/4.08	2	3.99/4.07	2
PRO	244	C	177.87	1		
THR	245	N	112.89	1	112.84	1
THR	245	HN	8.43	1	8.43	1
THR	245	CA	64.86	1	64.84	1
THR	245	HA	4.23	1	4.22	1
THR	245	CB	68.91	1	68.85	1
THR	245	HB	4.28	1	4.27	1
THR	245	CG2	22.14	1	22.00	1
THR	245	HG2	1.24	1	1.23	1
THR	245	C	176.05	1		
THR	246	N	109.19	1	109.14	1
THR	246	HN	7.65	1	7.63	1
THR	246	CA	61.90	1	61.73	1
THR	246	HA	4.43	1	4.43	1
THR	246	CB	71.36	1	71.13	1
THR	246	HB	4.28	1	4.27	1
THR	246	CG2	21.12	1	21.09	1
THR	246	HG2	1.20	1	1.19	1
THR	246	C	176.50	1		
GLY	247	N	111.01	1	110.98	1
GLY	247	HN	8.42	1	8.42	1
GLY	247	CA	45.77	1	45.67	1
GLY	247	HA2/HA3	3.77/4.23	2	3.77/4.23	2
GLY	247	C	173.96	1		
LYS	248	N	121.11	1	121.11	1
LYS	248	HN	7.93	1	7.93	1
LYS	248	CA	55.67	1	55.44	1
LYS	248	HA	4.43	1	4.42	1
LYS	248	CB	33.37	1	33.27	1
LYS	248	HB2/HB3	1.63/1.83	2	1.61/1.83	2

Residue name	Residue number	Atom name	Chemical shift [ppm]	Ambiguity	Chemical shift [ppm]	Ambiguity
			FIR RRM1-RRM2 free		FIR RRM1-RRM2 bound	
LYS	248	CG	24.98	1	24.87	1
LYS	248	HG2/HG3	1.36	2	1.35	2
LYS	248	CD	29.13	1	28.93	1
LYS	248	HD2/HD3	1.60	2	1.59	2
LYS	248	HE2/HE3	2.95	2	2.93	2
LYS	248	C	176.79	1		
HIS	249	N	120.09	1	120.10	1
HIS	249	HN	8.51	1	8.50	1
HIS	249	CA	55.55	1	55.42	1
HIS	249	HA	5.37	1	5.37	1
HIS	249	CB	32.12	1	32.11	1
HIS	249	HB2/HB3	3.28/2.59	2	3.27/2.58	2
HIS	249	CD2	119.78	1	119.73	1
HIS	249	HD2	7.07	1	7.07	1
HIS	249	CE1	139.39	1	139.20	1
HIS	249	HE1	7.42	1	7.42	1
HIS	249	C	176.57	1		
LYS	250	N	118.23	1	118.17	1
LYS	250	HN	8.61	1	8.62	1
LYS	250	CA	56.41	1	56.42	1
LYS	250	HA	4.43	1	4.42	1
LYS	250	CB	34.41	1	34.33	1
LYS	250	HB2/HB3	1.36/2.46	2	1.36/2.44	2
LYS	250	CG	25.96	1	25.83	1
LYS	250	HG2/HG3	1.35/1.48	2	1.35/1.48	2
LYS	250	CD	30.08	1	29.81	1
LYS	250	HD2/HD3	1.74/1.81	2	1.75/1.81	2
LYS	250	CE			41.63	1
LYS	250	HE2/HE3			3.01/2.95	2
LYS	250	C	177.99	1		
GLY	251	N	114.40	1	114.30	1
GLY	251	HN	11.23	1	11.21	1
GLY	251	CA	45.79	1	45.67	1
GLY	251	HA2/HA3	4.30/4.19	2	4.31/4.13	2
GLY	251	C	172.40	1		
TYR	252	N	112.78	1	112.77	1
TYR	252	HN	7.14	1	7.13	1
TYR	252	CA	54.66	1	54.64	1
TYR	252	HA	5.62	1	5.61	1
TYR	252	CB	41.28	1	41.28	1
TYR	252	HB2/HB3	2.76/3.02	2	2.76/3.01	2
TYR	252	CD1/CD2	133.71	3	133.74	3
TYR	252	HD1/HD2	6.81	3	6.82	3
TYR	252	CE1/CE2	118.05	3	118.02	3
TYR	252	HE1/HE2	6.81	3	6.81	3
TYR	252	C	172.69	1		
GLY	253	N	106.02	1	105.97	1
GLY	253	HN	8.81	1	8.77	1
GLY	253	CA	45.63	1	45.65	1
GLY	253	HA2/HA3	4.04/4.07	2	4.02/4.08	2
GLY	253	C	183.54	1		
PHE	254	N	113.11	1	113.01	1
PHE	254	HN	7.81	1	7.80	1
PHE	254	CA	56.51	1	56.35	1
PHE	254	HA	5.77	1	5.73	1
PHE	254	CB	44.55	1	44.51	1
PHE	254	HB2/HB3	2.76	2	2.73/2.75	2
PHE	254	CD1/CD2	132.30	3	132.27	3
PHE	254	HD1/HD2	7.27	3	7.26	3
PHE	254	CE1/CE2	130.94	3	131.05	3
PHE	254	HE1/HE2	7.33	3	7.34	3
PHE	254	C	175.87	1		
ILE	255	N	122.97	1	122.97	1
ILE	255	HN	8.73	1	8.77	1
ILE	255	CA	59.97	1	59.84	1
ILE	255	HA	4.59	1	4.55	1
ILE	255	CB	41.53	1	41.60	1
ILE	255	HB	1.23	1	1.23	1
ILE	255	CG1	28.34	1	28.34	1
ILE	255	HG12/HG13	1.05/0.59	2	1.05/0.58	2
ILE	255	CG2	17.94	1	18.18	1
ILE	255	HG2	0.23	1	0.22	1
ILE	255	CD1	14.23	1	14.22	1
ILE	255	HD1	-0.13	1	-0.13	1
ILE	255	C	173.90	1		
GLU	256	N	125.76	1	125.75	1
GLU	256	HN	8.78	1	8.83	1
GLU	256	CA	54.54	1	54.55	1

Residue name	Residue number	Atom name	Chemical shift [ppm]	Ambiguity	Chemical shift [ppm]	Ambiguity
			FIR RRM1-RRM2 free		FIR RRM1-RRM2 bound	
GLU	256	HA	5.04	1	5.03	1
GLU	256	CB	32.76	1	32.54	1
GLU	256	HB2/HB3	2.17/1.78	2	2.17/1.77	2
GLU	256	CG	36.57	1	36.55	1
GLU	256	HG2/HG3	2.28/2.16	2	2.23/2.16	2
GLU	256	C	175.96	1		
TYR	257	N	125.34	1	125.57	1
TYR	257	HN	8.84	1	8.83	1
TYR	257	CA	59.10	1	59.19	1
TYR	257	HA	4.95	1	4.85	1
TYR	257	CB	41.41	1	41.10	1
TYR	257	HB2/HB3	2.77/3.58	2	2.74/3.56	2
TYR	257	CD1/CD2	133.27	3	133.49	3
TYR	257	HD1/HD2	6.96	3	6.96	3
TYR	257	CE1/CE2	117.73	3	117.59	3
TYR	257	HE1/HE2	6.67	3	6.66	3
TYR	257	HH	8.80	1	8.77	1
TYR	257	C	176.01	1		
GLU	258	N	119.59	1	119.56	1
GLU	258	HN	8.27	1	8.29	1
GLU	258	CA	59.80	1	59.97	1
GLU	258	HA	4.11	1	4.08	1
GLU	258	CB	30.48	1	30.37	1
GLU	258	HB2/HB3	2.22/1.97	2	2.20/1.95	2
GLU	258	CG	36.90	1	36.92	1
GLU	258	HG2/HG3	2.24/2.14	2	2.21/2.12	2
GLU	258	C	176.91	1		
LYS	259	N	116.19	1	116.15	1
LYS	259	HN	8.76	1	8.80	1
LYS	259	CA	54.57	1	54.34	1
LYS	259	HA	4.74	1	4.73	1
LYS	259	CB	35.02	1	34.84	1
LYS	259	HB2/HB3	2.02/1.90	2	2.00/1.89	2
LYS	259	CG	25.12	1	25.02	1
LYS	259	HG2/HG3	1.58/1.49	2	1.58/1.48	2
LYS	259	CD	29.25	1	29.17	1
LYS	259	HD2/HD3	1.76	2	1.76	2
LYS	259	CE	42.29	1	42.02	1
LYS	259	HE2/HE3	3.05	2	3.04	2
ALA	260	CA	55.21	1	55.02	1
ALA	260	HA	3.94	1	3.91	1
ALA	260	CB	18.52	1	18.29	1
ALA	260	HB	1.44	1	1.42	1
GLN	261	CA	58.65	1	58.57	1
GLN	261	CB	28.26	1		
GLN	261	C	177.18	1		
SER	262	N	112.33	1	112.38	1
SER	262	HN	6.86	1	6.83	1
SER	262	CA	61.95	1	62.02	1
SER	262	HA	4.01	1	4.00	1
SER	262	CB	63.11	1	63.01	1
SER	262	HB2/HB3	3.48/2.78	2	3.48/2.86	2
SER	262	C	174.15	1		
SER	263	N	113.62	1	113.58	1
SER	263	HN	6.93	1	6.88	1
SER	263	CA	61.17	1	60.96	1
SER	263	HA	3.72	1	3.68	1
SER	263	CB	62.88	1	62.63	1
SER	263	HB2/HB3	3.84/3.74	2	3.81/3.76	2
SER	263	C	175.48	1		
GLN	264	N	121.16	1	121.53	1
GLN	264	HN	7.31	1	7.40	1
GLN	264	CA	58.67	1	58.62	1
GLN	264	HA	3.94	1	3.93	1
GLN	264	CB	28.32	1	28.26	1
GLN	264	HB2/HB3	2.09	2	2.09	2
GLN	264	CG	33.73	1	33.70	1
GLN	264	HG2/HG3	2.38	2	2.36	2
GLN	264	C	171.68			
ASP	265	N	121.20	1	121.48	1
ASP	265	HN	8.20	1	8.32	1
ASP	265	CA	57.04	1	57.17	1
ASP	265	HA	4.26	1	4.24	1
ASP	265	CB	40.31	1	40.48	1
ASP	265	HB2/HB3	2.75/2.72	2	2.74/2.72	2
ASP	265	C	178.07	1		
ALA	266	N	122.41	1	122.67	1
ALA	266	HN	7.94	1	7.92	1

Residue name	Residue number	Atom name	Chemical shift [ppm]	Ambiguity	Chemical shift [ppm]	Ambiguity
			FIR RRM1-RRM2 free		FIR RRM1-RRM2 bound	
ALA	266	CA	55.00	1	55.10	1
ALA	266	HA	2.34	1	2.40	1
ALA	266	CB	19.70	1	19.26	1
ALA	266	HB	1.34	1	1.25	1
ALA	266	C	179.51	1		
VAL	267	N	116.40	1	116.66	1
VAL	267	HN	7.55	1	7.71	1
VAL	267	CA	66.44	1	66.14	1
VAL	267	HA	3.44	1	3.47	1
VAL	267	CB	31.59	1	31.50	1
VAL	267	HB	2.06	1	2.06	1
VAL	267	CG1/CG2	21.88/23.41	2	21.82/23.58	2
VAL	267	HG1/HG2	0.97/1.01	2	0.96/0.99	2
VAL	267	C	177.72	1		
SER	268	N	112.11	1	112.13	1
SER	268	HN	7.65	1	7.74	1
SER	268	CA	61.03	1	60.75	1
SER	268	HA	4.20	1	4.21	1
SER	268	CB	63.47	1	63.55	1
SER	268	HB2/HB3	3.94	2	3.95/3.93	2
SER	268	C	176.28	1		
SER	269	N	112.86	1	112.55	1
SER	269	HN	7.50	1	7.36	1
SER	269	CA	60.98	1	60.48	1
SER	269	HA	4.58	1	4.57	1
SER	269	CB	65.21	1	65.61	1
SER	269	HB2/HB3	3.75/3.49	2	3.70/3.52	2
SER	269	C	175.61	1		
MET	270	N	115.02	1	115.41	1
MET	270	HN	7.85	1	7.75	1
MET	270	CA	54.05	1	54.34	1
MET	270	HA	4.88	1	4.94	1
MET	270	CB	32.03	1	31.63	1
MET	270	HB2/HB3	1.85/1.98	2	1.94/2.07	2
MET	270	CG	33.48	1	33.81	1
MET	270	HG2/HG3	1.64/1.86	2	1.61/1.96	2
MET	270	CE	16.05	1	15.35	1
MET	270	HE	0.60	1	0.19	1
MET	270	C	176.76	1		
ASN	271	N	117.81	1	118.36	1
ASN	271	HN	7.50	1	7.83	1
ASN	271	CA	56.75	1	57.06	1
ASN	271	HA	4.44	1	4.46	1
ASN	271	CB	39.08	1	39.28	1
ASN	271	HB2/HB3	2.86/2.97	2	2.83/2.91	2
ASN	271	ND2	113.21	1	113.40	1
ASN	271	HD21/HD22	7.84/6.43	2	7.89/6.54	2
ASN	271	C	175.90	1		
LEU	272	N	123.66	1	123.40	1
LEU	272	HN	9.17	1	9.23	1
LEU	272	CA	57.24	1	57.16	1
LEU	272	HA	3.90	1	3.88	1
LEU	272	CB	38.48	1	38.40	1
LEU	272	HB2/HB3	1.54/2.03	2	1.54/2.03	2
LEU	272	CG	27.15	1	27.11	1
LEU	272	HG	1.51	1	1.52	1
LEU	272	CD1/CD2	24.94/22.23	2	24.98/22.18	2
LEU	272	HD1/HD2	0.81/0.71	2	0.80/0.70	2
LEU	272	C	174.43	1		
PHE	273	N	121.19	1	120.99	1
PHE	273	HN	7.71	1	7.59	1
PHE	273	CA	58.71	1	58.72	1
PHE	273	HA	4.36	1	4.29	1
PHE	273	CB	41.71	1	41.47	1
PHE	273	HB2/HB3	3.16/3.01	2	3.31/2.97	2
PHE	273	CD1/CD2	132.34	3	132.84	3
PHE	273	HD1/HD2	7.11	3	7.17	3
PHE	273	CE1/CE2	130.98	3	130.29	3
PHE	273	HE1/HE2	7.26	3	6.92	3
PHE	273	C	175.06	1		
ASP	274	N	126.85	1	127.06	1
ASP	274	HN	7.73	1	7.91	1
ASP	274	CA	53.60	1	53.70	1
ASP	274	HA	4.34	1	4.31	1
ASP	274	CB	40.71	1	40.72	1
ASP	274	HB2/HB3	2.10/2.45	2	2.21/2.46	2
ASP	274	C	173.99	1		
LEU	275	N	128.64	1	128.83	1

Residue name	Residue number	Atom name	Chemical shift [ppm]	Ambiguity	Chemical shift [ppm]	Ambiguity
			FIR RRM1-RRM2 free		FIR RRM1-RRM2 bound	
LEU	275	HN	8.35	1	8.49	1
LEU	275	CA	53.91	1	53.58	1
LEU	275	HA	4.71	1	4.77	1
LEU	275	CB	44.47	1	43.88	1
LEU	275	HB2/HB3	1.46/2.01	2	1.46/2.09	2
LEU	275	CG	27.64	1	27.54	1
LEU	275	HG	1.57	1	1.58	1
LEU	275	CD1/CD2	24.02/27.33	2	24.38/27.14	2
LEU	275	HD1/HD2	1.09/0.98	2	1.04/0.97	2
LEU	275	C	176.49	1		
GLY	276	N	114.43	1	114.25	1
GLY	276	HN	8.85	1	9.06	1
GLY	276	CA	46.44	1	46.37	1
GLY	276	HA2/HA3	3.66	2	3.67/3.62	2
GLY	276	C	174.29	1		
GLY	277	N	106.30	1	106.93	1
GLY	277	HN	8.37	1	8.46	1
GLY	277	CA	45.07	1	45.01	1
GLY	277	HA2/HA3	3.60/4.07	2	3.57/4.06	2
GLY	277	C	174.04	1		
GLN	278	N	117.95	1	117.60	1
GLN	278	HN	7.21	1	7.11	1
GLN	278	CA	53.94	1	53.74	1
GLN	278	HA	4.38	1	4.33	1
GLN	278	CB	32.17	1	31.81	1
GLN	278	HB2/HB3	1.99/2.13	2	1.94/2.13	2
GLN	278	CG	34.48	1	34.25	1
GLN	278	HG2/HG3	2.28/2.37	2	2.25/2.33	2
GLN	278	NE2	109.67	1	109.78	1
GLN	278	HE21/HE22	6.92/6.79	2	6.91/6.73	2
GLN	278	C	173.13	1		
TYR	279	N	120.05	1	120.84	1
TYR	279	HN	8.15	1	8.24	1
TYR	279	CA	56.60	1	55.97	1
TYR	279	HA	4.78	1	4.81	1
TYR	279	CB	38.12	1	38.06	1
TYR	279	HB2/HB3	2.56/2.74	2	2.54/2.73	2
TYR	279	CD1/CD2	132.81	3	132.72	3
TYR	279	HD1/HD2	6.99	3	6.98	3
TYR	279	CE1/CE2	118.42	3	118.26	3
TYR	279	HE1/HE2	6.77	3	6.76	3
TYR	279	C	175.87	1		
LEU	280	N	123.80	1	123.46	1
LEU	280	HN	9.16	1	9.24	1
LEU	280	CA	55.61	1	55.32	1
LEU	280	HA	4.46	1	4.47	1
LEU	280	CB	43.59	1	43.52	1
LEU	280	HB2/HB3	1.37/2.31	2	1.35/2.26	2
LEU	280	CG	26.61	1	26.64	1
LEU	280	HG	2.24	1	2.25	1
LEU	280	CD1/CD2	26.85/23.25	2	26.89/23.30	2
LEU	280	HD1/HD2	1.06/0.87	2	1.04/0.87	2
LEU	280	C	177.27	1		
ARG	281	N	119.40	1	119.43	1
ARG	281	HN	7.76	1	7.76	1
ARG	281	CA	52.83	1	52.62	1
ARG	281	HA	5.25	1	5.27	1
ARG	281	CB	31.42	1	31.42	1
ARG	281	HB2/HB3	1.76	2	1.74	2
ARG	281	CG	27.13	1	27.26	1
ARG	281	HG2/HG3	1.54/1.02	2	1.53/0.98	2
ARG	281	CD	42.40	1	42.22	1
ARG	281	HD2/HD3	3.18/3.12	2	3.16/3.11	2
ARG	281	C	174.71	1		
VAL	282	N	122.11	1	121.52	1
VAL	282	HN	8.76	1	8.68	1
VAL	282	CA	60.17	1	59.98	1
VAL	282	HA	5.39	1	5.37	1
VAL	282	CB	36.16	1	36.16	1
VAL	282	HB	1.81	1	1.78	1
VAL	282	CG1/CG2	23.07/23.46	2	22.97	2
VAL	282	HG1/HG2	1.01/1.04	2	0.99	2
VAL	282	C	174.41	1		
GLY	283	N	110.74	1	110.40	1
GLY	283	HN	9.05	1	9.02	1
GLY	283	CA	45.09	1	45.03	1
GLY	283	HA2/HA3	4.60/4.03	2	4.57/4.03	2
GLY	283	C	171.89	1		

Residue name	Residue number	Atom name	Chemical shift [ppm]	Ambiguity	Chemical shift [ppm]	Ambiguity
			FIR RRM1-RRM2 free		FIR RRM1-RRM2 bound	
LYS	284	N	119.20	1	119.29	1
LYS	284	HN	8.60	1	8.60	1
LYS	284	CA	57.01	1	56.94	1
LYS	284	HA	4.44	1	4.41	1
LYS	284	CB	33.19	1	33.24	1
LYS	284	HB2/HB3	1.85/2.00	2	1.83/1.98	2
LYS	284	CG	25.32	1	25.26	1
LYS	284	HG2/HG3	1.54	2	1.54	2
LYS	284	CD	29.53	1	29.41	1
LYS	284	HD2/HD3	1.74	2	1.73	2
LYS	284	CE	42.16	1	42.03	1
LYS	284	HE2/HE3	3.00	2	2.99	2
LYS	284	C	176.34	1		
ALA	285	N	121.04	1	120.97	1
ALA	285	HN	7.76	1	7.71	1
ALA	285	CA	52.56	1	52.41	1
ALA	285	HA	4.75	1	4.75	1
ALA	285	CB	19.69	1	19.53	1
ALA	285	HB	1.53	1	1.52	1
ALA	285	C	178.18	1		
VAL	286	N	114.62	1	114.72	1
VAL	286	HN	9.84	1	9.85	1
VAL	286	CA	60.76	1	60.41	1
VAL	286	HA	4.52	1	4.50	1
VAL	286	CB	31.94	1	31.78	1
VAL	286	HB	2.38	1	2.36	1
VAL	286	CG1/CG2	21.59/19.06	2	21.42/19.07	2
VAL	286	HG1/HG2	0.79/0.56	2	0.78/0.55	2
VAL	286	C	173.13	1		
THR	287	N	108.36	1	108.29	1
THR	287	HN	7.24	1	7.22	1
THR	287	CA	58.08	1	57.87	1
THR	287	HA	4.67	1	4.65	1
THR	287	CB	70.25	1	70.01	1
THR	287	HB	4.21	1	4.20	1
THR	287	CG2	23.57	1	23.04	1
THR	287	HG2	1.09	1	1.08	1
PRO	288	CA	61.54	1	61.50	1
PRO	288	HA	4.36	1	4.34	1
PRO	288	CB	30.43	1	30.27	1
PRO	288	HB2/HB3	2.00/1.88	2	2.06/1.87	2
PRO	288	CG	26.18	1	26.13	1
PRO	288	HG2/HG3	2.32/1.76	2	2.31/1.75	2
PRO	288	CD	50.18	1	50.20	1
PRO	288	HD2/HD3	3.54/3.93	2	3.53/3.92	2
PRO	291	CA	63.71	1	63.63	1
LEU	292	N	120.02	1	120.10	1
LEU	292	HN	8.18	1	8.17	1
LEU	292	CA	55.00	1		
LEU	292	CB	41.91	1		
LEU	293	N	121.89	1	121.69	1
LEU	293	HN	7.87	1	7.85	1
LEU	293	CA	55.02	1	55.01	1
LEU	293	CB	42.47	1		
THR	294	N	116.94	1	116.96	1
THR	294	HN	7.95	1	7.95	1
THR	294	CA	59.62	1	59.51	1
THR	294	CB	69.83	1		
PRO	295	CA	63.28	1	63.21	1
PRO	295	CB	32.21	1		
ALA	296	N	125.00	1	124.97	1
ALA	296	HN	8.32	1	8.30	1
ALA	296	CA	52.63	1	52.54	1
ALA	296	HA	4.37	1	4.42	1
ALA	296	CB	19.38	1	19.41	1
ALA	296	HB	1.40	1	1.46	1
ALA	296	C	176.99	1		
THR	297	N	118.44	1	118.41	1
THR	297	HN	7.64	1	7.66	1
THR	297	CA	63.05	1	62.97	1
THR	297	CB	70.90	1		

Residue name	Residue number	Atom name	Chemical shift [ppm]	Ambiguity	Chemical shift [ppm]	Ambiguity
			FBP Nbox free		FBP Nbox bound	
GLY	22	CA	43.42	1		
ALA	23	N	123.85	1		
ALA	23	HN	8.59	1		
ALA	23	CA	52.73	1		
ALA	23	HA	4.37	1		
ALA	23	CB	19.46	1		
ALA	23	HB	1.39	1		
MET	24	N	119.74	1		
MET	24	HN	8.47	1		
MET	24	CA	55.65	1	55.59	1
MET	24	HA	4.47	1	4.49	1
MET	24	CB	32.92	1	32.56	1
MET	24	HB2/HB3	2.06/1.99	2	2.07/2.00	2
MET	24	CG	32.00	1	32.47	1
MET	24	HG2/HG3	2.60/2.55	2	2.61/2.55	2
MET	24	CE			17.05	1
MET	24	HE			2.10	1
GLY	25	N	109.93	1		
GLY	25	HN	8.32	1		
GLY	25	CA	45.36	1	45.35	1
GLY	25	HA2/HA3	3.90	2	3.91	2
TYR	26	N	120.22	1		
TYR	26	HN	8.02	1		
TYR	26	CA	57.93	1	58.07	1
TYR	26	HA	4.59	1	4.59	1
TYR	26	CB	39.00	1	38.65	1
TYR	26	HB2/HB3	3.01/2.96	2	3.03	2
TYR	26	HD1/HD2			7.12	3
TYR	26	HE1/HE2			6.85	3
VAL	27	N	122.41	1		
VAL	27	HN	8.00	1		
VAL	27	CA	62.29	1	62.78	1
VAL	27	HA	4.04	1	3.99	1
VAL	27	CB	33.01	1	32.70	1
VAL	27	HB	2.00	1	2.02	1
VAL	27	CG1/CG2	20.81	2	21.06	2
VAL	27	HG1/HG2	0.88	2	0.89	2
ASN	28	N	122.08	1		
ASN	28	HN	8.32	1		
ASN	28	CA	53.40	1	53.86	1
ASN	28	HA	4.65	1	4.65	1
ASN	28	CB	39.16	1	38.98	1
ASN	28	HB2/HB3	2.83/2.77	2	2.90	2
ASP	29	N	121.69	1		
ASP	29	HN	8.29	1		
ASP	29	CA	54.87	1	56.61	1
ASP	29	HA	4.54	1	4.48	1
ASP	29	CB	41.25	1	40.92	1
ASP	29	HB2/HB3	2.64	2	2.75/2.67	2
ALA	30	N	123.63	1		
ALA	30	HN	8.19	1		
ALA	30	CA	53.68	1	54.77	1
ALA	30	HA	4.17	1	4.10	1
ALA	30	CB	18.79	1	18.79	1
ALA	30	HB	1.27	1	1.20	1
PHE	31	N	118.11	1		
PHE	31	HN	8.08	1		
PHE	31	CA	58.46	1	61.04	1
PHE	31	HA	4.54	1	4.24	1
PHE	31	CB	39.04	1	39.57	1
PHE	31	HB2/HB3	3.22/3.08	2	3.22	2
PHE	31	HD1/HD2			7.22	3
PHE	31	HE1/HE2			7.52	3
LYS	32	N	118.11	1		
LYS	32	HN	7.95	1		
LYS	32	CA	57.83	1	59.78	1
LYS	32	HA	4.12	1	3.84	1
LYS	32	CB	32.85	1	32.15	1
LYS	32	HB2/HB3	1.81	2	1.92	1
LYS	32	CG	24.78	1	25.43	1
LYS	32	HG2/HG3	1.44/1.38	2	1.47/1.67	2
LYS	32	CD	29.19	1	29.28	1
LYS	32	HD2/HD3	1.68	2	1.73	2
LYS	32	CE			41.88	1
LYS	32	HE2/HE3	3.00	2	3.01	2
ASP	33	N	120.12	1		
ASP	33	HN	8.17	1		
ASP	33	CA	55.32	1	56.74	1
ASP	33	HA	4.52	1	4.47	1

Residue name	Residue number	Atom name	Chemical shift [ppm]	Ambiguity	Chemical shift [ppm]	Ambiguity
			FBP Nbox free		FBP Nbox bound	
ASP	33	CB	41.28	1	41.13	1
ASP	33	HB2/HB3	2.70	2	2.79/2.76	2
ALA	34	N	123.28	1		
ALA	34	HN	8.09	1		
ALA	34	CA	54.35	1	54.53	1
ALA	34	HA	4.14	1	3.92	1
ALA	34	CB	18.75	1		
ALA	34	HB	1.45	1	1.30	1
LEU	35	N	119.46	1		
LEU	35	HN	8.04	1		
LEU	35	CA	56.69	1	58.45	1
LEU	35	HA	4.20	1	3.95	1
LEU	35	CB	41.90	1	42.19	1
LEU	35	HB2/HB3	1.73/1.60	2	1.71/1.47	2
LEU	35	CG	27.08	1	26.46	1
LEU	35	HG	1.62	1	1.34	1
LEU	35	CD1/CD2	24.57/23.68	2	25.45/24.19	2
LEU	35	HD1/HD2	0.90/0.86	2	0.83/0.78	2
GLN	36	N	119.12	1		
GLN	36	HN	8.04	1		
GLN	36	CA	57.39	1	59.04	1
GLN	36	HA	4.16	1	4.07	1
GLN	36	CB	28.70	1	28.02	1
GLN	36	HB2/HB3	2.13/2.10	2	2.27/2.29	2
GLN	36	CG	33.87	1	33.63	1
GLN	36	HG2/HG3	2.43/2.38	2	2.49/2.53	2
ARG	37	N	120.16	1		
ARG	37	HN	8.08	1		
ARG	37	CA	57.61	1		
ARG	37	HA	4.17	1		
ARG	37	CB	30.47	1		
ARG	37	HB2/HB3	1.83	2	1.89	2
ARG	37	CG	27.43	1		
ARG	37	HG2/HG3	1.65	2	1.65/1.88	2
ARG	37	CD	43.42	1	43.57	1
ARG	37	HD2/HD3	3.19	2	2.95/3.47	2
ALA	38	N	122.58	1		
ALA	38	HN	8.00	1		
ALA	38	CA	53.53	1	55.26	1
ALA	38	HA	4.24	1	3.97	1
ALA	38	CB	18.78	1		
ALA	38	HB	1.45	1	1.62	1
ARG	39	N	118.79	1		
ARG	39	HN	8.02	1		
ARG	39	CA	57.10	1	59.09	1
ARG	39	HA	4.23	1	4.06	1
ARG	39	CB	30.63	1	29.82	1
ARG	39	HB2/HB3	1.88/1.83	2	2.00/1.94	2
ARG	39	CG	27.24	1	27.86	1
ARG	39	HG2/HG3	1.66	2	1.91/1.69	2
ARG	39	CD	43.54	1	43.47	1
ARG	39	HD2/HD3	3.20	2	3.19	2
GLN	40	N	120.57	1		
GLN	40	HN	8.14	1		
GLN	40	CA	56.66	1	58.50	1
GLN	40	HA	4.29	1	4.14	1
GLN	40	CB	29.22	1	28.42	1
GLN	40	HB2/HB3	2.11/2.06	2	2.21/2.26	2
GLN	40	CG	33.92	1	34.00	1
GLN	40	HG2/HG3	2.41/2.37	2	2.44/2.55	2
ILE	41	N	121.18	1		
ILE	41	HN	8.04	1		
ILE	41	CA	61.87	1	63.73	1
ILE	41	HA	4.09	1	3.87	1
ILE	41	CB	38.73	1	37.99	1
ILE	41	HB	1.88	1	1.98	1
ILE	41	CG1	27.45	1	28.69	1
ILE	41	HG12/HG13	1.50/1.20	2	1.70/1.37	2
ILE	41	CG2	17.42	1	18.65	1
ILE	41	HG2	0.92	1	0.96	1
ILE	41	CD1	12.89	1	13.56	1
ILE	41	HD1	0.86	1	0.98	1
ALA	42	N	126.53	1		
ALA	42	HN	8.14	1		
ALA	42	CA	52.82	1	53.97	1
ALA	42	HA	4.29	1	4.17	1
ALA	42	CB	19.16	1	18.40	1
ALA	42	HB	1.40	1	1.48	1
ALA	43	N	122.60	1		

Residue name	Residue number	Atom name	Chemical shift [ppm]	Ambiguity	Chemical shift [ppm]	Ambiguity
			FBP Nbox free		FBP Nbox bound	
ALA	43	HN	8.04	1		
ALA	43	CA	52.74	1	53.98	1
ALA	43	HA	4.28	1	4.17	1
ALA	43	CB	19.27	1	18.44	1
ALA	43	HB	1.40	1	1.48	1
LYS	44	N	120.19	1		
LYS	44	HN	8.10	1		
LYS	44	CA	56.39	1	57.47	1
LYS	44	HA	4.32	1	4.23	1
LYS	44	CB	32.93	1	32.45	1
LYS	44	HB2/HB3	1.84/1.77	2	1.95	1
LYS	44	CG	24.75	1	24.85	1
LYS	44	HG2/HG3	1.45/1.39	2	1.51/1.55	2
LYS	44	CD	29.17	1	29.34	1
LYS	44	HD2/HD3	1.68	2	1.73	2
LYS	44	CE			41.81	1
LYS	44	HE2/HE3	3.00	2	3.00	2
ILE	45	N	122.14	1		
ILE	45	HN	8.15	1		
ILE	45	CA	61.40	1	62.18	1
ILE	45	HA	4.19	1	4.10	1
ILE	45	CB	38.76	1	38.18	1
ILE	45	HB	1.88	1	1.90	1
ILE	45	CG1	27.39	1	27.51	1
ILE	45	HG12/HG13	1.50/1.20	2	1.46/1.09	2
ILE	45	CG2	17.42	1	17.38	1
ILE	45	HG2	0.92	1	0.79	1
ILE	45	CD1	12.87	1		
ILE	45	HD1	0.86	1		
GLY	46	N	112.86	1		
GLY	46	HN	8.46	1		
GLY	46	CA	45.43	1		
GLY	46	HA2/HA3	3.96/4.05	2		
GLY	47	N	108.74	1		
GLY	47	HN	8.22	1		
GLY	47	CA	45.43	1		
GLY	47	HA2/HA3	3.98	2		
ASP	48	N	120.40	1		
ASP	48	HN	8.29	1		
ASP	48	CA	54.42	1	54.27	1
ASP	48	HA	4.62	1	4.63	1
ASP	48	CB	41.48	1	41.54	1
ASP	48	HB2/HB3	2.65/2.69	2	2.63/2.70	2
ALA	49	N	124.29	1		
ALA	49	HN	8.31	1		
ALA	49	CA	52.93	1	52.68	1
ALA	49	HA	4.34	1	4.35	1
ALA	49	CB	19.26	1	19.37	1
ALA	49	HB	1.42	1	1.41	1
GLY	50	N	107.80	1		
GLY	50	HN	8.39	1		
GLY	50	CA	45.58	1	45.40	1
GLY	50	HA2/HA3	4.03	2	4.01	2
THR	51	N	112.90	1		
THR	51	HN	7.98	1		
THR	51	CA	61.73	1	61.60	1
THR	51	HA	4.44	1	4.45	1
THR	51	CB	70.18	1	70.07	1
THR	51	HB	4.32	1	4.32	1
THR	51	CG2	21.39	1	21.49	1
THR	51	HG2	1.20	1	1.21	1
SER	52	N	123.42	1		
SER	52	HN	8.00	1		
SER	52	CA	60.19	1	60.11	1
SER	52	HA	4.31	1	4.31	1
SER	52	CB	65.06	1	64.83	1
SER	52	HB2/HB3	3.87	2	3.87	2

Appendix VI Inter-molecular restraints used during structure calculations of FIR RRM1-RRM2 – FBP Nbox complex.

No.	FIR RRM1-RRM2 atom(s)	FBP Nbox peptide atom(s)	Upper bound [Å]
1	F232 HA	A30 QB	3.5
2	V228 QG1	L35 QQD	3.5
3	A231 QB	A30 QB	5.0
4	A231 QB	D33 HB2	3.5
5	A231 QB	D33 HB3	3.5
6	D224 HB2	I41 QG2/I41 QD1	5.0
7	D224 HB3	I41 QG2/I41 QD1	5.0
8	V228 QG1	I41 QG2/I41 QD1	2.8
9	V228 QG2	I41 QG2/I41 QD1	5.0
10	L220 QD1	I41 QG2/I41 QD1	3.5
11	L220 QD2	I41 QG2/I41 QD1	3.5
12	L275 QD1	I41 QG2/I41 QD1	2.8
13	S269 HB2	V27 QQG	5.0
14	S269 HB3	V27 QQG	5.0
15	S269 HA	V27 QQG	5.0
16	S227 HB2	L35 QQD	5.0
17	S227 HB3	L35 QQD	5.0
18	D265 HA	V27 QQG	5.0
19	D265 HB2	V27 QQG	5.0
20	D265 HB3	V27 QQG	5.0
21	L220 QD1	R39 HG2	5.0
22	L220 QD1	R39 HG3	5.0
23	A231 QB	R37 HD2	6.0
24	A231 QB	R37 HD3	6.0
25	S269 HB2	F31 QD	5.0
26	S269 HB3	F31 QD	5.0
27	S227 HB2	I41 QG2/I41 QD1	5.0
28	S227 HB3	I41 QG2/I41 QD1	5.0

No.	FIR RRM1-RRM2 atom(s)	FBP Nbox peptide atom(s)	Upper bound [Å]
29	L275 QD2	I41 QG2/I41 QD1	3.5
30	V228 QG1	A34 HA/L35 HA/A38 HA	2.8
31	V228 QG1	R37 HB2/ R37 HB3/I41 HB	3.5
32	V228 QG1	A38 QB	3.5
33	V228 QG1	I41 HG12	5.0
34	V228 QG2	A34 HA/L35 HA	5.0
35	V228 QG2	R37 HB2/ R37 HB3/I41 HB	5.0
36	V228 QG2	A38 QB	3.5
37	V228 QG2	A34 QB	5.0
38	A231 QB	A30 HA/A34 HA	2.8
39	A231 QB	F31 QB	6.0
40	A231 QB	R37 HB2	5.0
41	A231 QB	R37 HB3	5.0
42	A231 QB	A34 QB	5.0
43	F232 HA	A34 QB	5.0
44	L275 QD1	I41 HB	5.0
45	L275 QD1	A38 QB	3.5
46	F232 QD	A30 QB	5.0
47	F232 QE	A30 QB	5.0
48	F273 QE	I41 QG2	5.0
49	F273 QE	I41 QD1	5.0
50	F273 QD	I41 QD1	5.0
51	F232 QE	A34 HA	3.5
52	F273 QD	A38 HA	5.0
53	F273 QE	A38 HA	3.5
54	F273 QD	L35 HG	5.0
55	F232 QD	V27 QQG	5.0
56	F232 QE	V27 QQG	5.0
57	F232 QD	F31 HA	2.8
58	V228 QG1	F31 QD	3.5
59	V228 QG1	F31 QE	3.5

No.	FIR RRM1-RRM2 atom(s)	FBP Nbox peptide atom(s)	Upper bound [Å]
60	L220 QD1	K44 QE	5.0
61	D219 HB2	I45 HA	5.0
62	D219 HB3	I45 HA	5.0
63	L220 QD1	I41 HB	5.0
64	L220 QD1	A38 QB	5.0
65	V228 HA	R37 HB2/R37 HB3	5.0
66	A231 QB	D33 HA	5.0
67	A231 QB	R37 QG	5.0
68	M270 QE	A38 QB	6.0
69	M270 QE	A34 QB	6.0
70	L275 QD1	A38 HA/L35 HA/R39 HA	3.5
71	L275 QD1	A42 QB	5.0
72	D224 HA	I41 QG2/I41 QD1	3.5